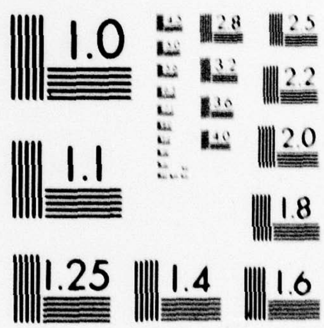
AD-A073 574 PURDUE UNIV LAFAYETTE IND DEPT OF STATISTICS F/G 12/1
SOME CONTRIBUTIONS TO GAMMA-MINIMAX AND EMPIRICAL BAYES SELECTI--ETC(U)
AUG 79 P HSIAO N00014-75-C-0455
UNCLASSIFIED MMS-79-16 NL

1 OF 2
AD
A073574



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12) LEVEL II

PURDUE UNIVERSITY

DEPARTMENT OF STATISTICS

DIVISION OF MATHEMATICAL SCIENCES

79 09 5 023

SOME CONTRIBUTIONS TO GAMMA-MINIMAX AND EMPIRICAL BAYES SELECTION PROCEDURES

by

Ping Hsiao
Purdue University

See 1473 in back

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #79-16

August 1979

*This research was supported by the Office of Naval Research contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

INTRODUCTION

Until about 1950's, the statistical inference problems were primarily formulated as problems of the estimation of parameters and tests of hypotheses. Estimation problems, in general, are decision problems with infinitely many actions whereas hypotheses testing problems are two actions problems. For problems of comparing k populations ($k > 2$), usually, more than 2 actions should be considered. Thus it is not quite realistic to treat them only as hypotheses testing problems. The classic tests of homogeneity were found to be inadequate in two respects. First, the formulation is not designed to answer many questions which are of real interest to the experimenter. Second, we almost always reject the null hypothesis which says all the parameters are equal if enough data are collected. To eliminate the shortcomings, one should formulate the problems as multiple decision problems. Mosteller (1948), Paulson (1949), Bahadur (1950) and Bahadur and Robbins (1950) were among the earlier researchers to do so, thus laying the groundwork for the investigation of selection and ranking procedures.

'Indifference zone' approach, proposed by Bechhofer (1954) is one of the two basic formulations for ranking and selection problems. In this approach, a single population (or a fixed

size of populations) is selected and is guaranteed to be the one of interest with probability P^* if the parameters lie outside some subset, the zone of indifference. Another basic formulation, which is due to Gupta (1956, 1963, 1965), is the 'subset selection' approach. In this approach, one wishes to select a subset which contains the population (or populations) of interest with a minimum probability P^* over the whole parameter space. The size of the selected subset depends on the outcome of the experiment and is not fixed in advance. Using these two approaches, a large number of contributions have been made. A complete bibliography can be found in a forthcoming monograph of Gupta and Panchapakesan (1979).

Bayes approach for selection and ranking problems has also been considered. Recent contributions made in this framework are Hsu (1977), Gupta and Hsu (1978), Miescke (1978) and Kim (1979). Bayesian analysis is attractive if a prior distribution for the unknown parameters can be specified exactly. However, it is often that one can only have partial prior information. In this case, the prior is restricted in some sub-class Γ of all prior distributions. The Γ-minimax criterion then requires the use of decisions which minimize the maximum Bayes risk over Γ. Such a principle has been used in multiple decision problems by Randles and Holland (1971), Gupta and Huang (1975, 1977), Berger (1977), Miescke (1979) and Kim (1979). The first two chapters in this thesis are related to Γ-minimax rules.

There are situations where a statistical decision problem occurs repeatedly and independently. Then frequently, empirical

Bayes approach becomes appropriate for consideration. In this approach, one assumes no prior information about the parameters except for the existence of a prior distribution τ. By use of this empirical Bayes approach, one can then guarantee that the rules one uses are almost as good as the Bayes rule with respect to τ for large samples. Empirical Bayes rules for multiple decision problems have been derived by Deely (1965), Van Ryzin (1970), Huang (1975). Van Ryzin and Susarla (1977) and Singh (1977).

Besides the comparison of k populations among themselves, sometimes, in practice, one wishes to compare them with a control population (or a standard population). There are many situations where one wants to select populations better than a control. But there are other cases where one is interested in selecting populations close to a control. Contributions related to these topics can be found in Chapter 20 of Gupta and Panchapakesan (1979).

In this thesis, some results about the Γ-minimax rules and empirical Bayes rules have been obtained. In Chapter I, a problem of selecting populations close to a control is considered. Under the assumption that populations are normally distributed, our goal is to select the 'good' populations. A '0-1' type loss is introduced. When the control parameter is known, we derive a Γ-minimax rule. When it is unknown, a restricted Γ-minimax rule is derived. We also find Bayes rules and minimax rules for the unknown parameter case. A comparison among these three rules is made. For Γ-minimax rules, we show some optimal properties and

some general distributions for which Γ-minimax rules can be found.

The problem of selecting the t-best populations is discussed in Chapter II. It is shown that if the populations have PF_2 densities, then the natural selection rule - which selects the populations with the largest t sample values - is a Γ-minimax rule. This result has also been extended to the case where the populations are not necessarily independent. Also, by a simultaneous selection of the t-best populations for all $1 \leq t \leq k-1$, a Γ-minimax rule for the complete ranking of k populations is derived.

Chapter III deals with a problem of selecting populations which are 'better' than a control. Under a linear loss, we derive a sequence of empirical Bayes rules for uniformly distributed populations. When the priors are assumed to have bounded supports, empirical Bayes rules are obtained for more general distributions. Based on Monte Carlo studies, tables are computed for the smallest sample size required for the empirical Bayes rules to be 'close' to the true Bayes rules.

CHAPTER I

Γ-MINIMAX PROCEDURE FOR SELECTING POPULATIONS CLOSE TO A CONTROL

1.1 Introduction

Problems of selecting populations close to a control arise frequently in inductrial production such as to match parts or to imitate some popular goods in the market. This may be the first step for quality control, since after knowing the "good" populations, we may find ways to improve production so that all products are "exactly" alike to a fairly good degree of precision. Thus the selection problem is interesting and challenging.

Many authors have considered the problem of comparing populations with a control under different types of formulations. Paulson (1952), Bechhofer and Turnbull (1974) discussed problems of selecting the best population if the best population is better than the control. Dunnett (1955), Gupta and Sobel (1958) considered the problem of selecting a subset containing all populations better than the control. Lehmann (1961), Randles and Hollander (1971) dealt with the problem of selecting populations better than a control. Bhattacharyya (1956, 1958), Tong (1969), Seeger (1972), Huang (1975), and Kim (1979) have considered partitioning a set of populations with respect to a control. Non-parametric procedures related to some aspect of the problem have been studied by Rizvi, Sobel and Woodworth (1968), Puri and Puri (1969). However, very few papers

have been devoted to the discussion of selecting populations close to a control. A. K. Singh (1977) considered this problem and derived Bayes rules and empirical Bayes rules for Poisson, Geometric and Binomial populations. Except in rare situations, information concerning the prior distribution of a parameter is likely to be incomplete. Hence the use of Bayes rules is hard to justify. The use of partial or incomplete prior information in statistical inference has led to the development of the 'so-called' Γ-minimax criterion, a term initially employed by Blum and Rosenblatt (1967). The original idea of Γ-minimaxity is due to Robbins(1951). To be more precise, although we may not know the prior distribution completely we may have enough information to specify that the prior is a member of a subset Γ of the class of all priors. Γ-minimax criterion then requires one to use the decision rule which minimizes the maximum expected risk over Γ. It is interesting to note that if Γ contains only a single prior, then the Γ-minimax rule is just the Bayes rule for that prior. At the other extreme, when Γ consists of all priors, the Γ-minimax rule reduces to the minimax rule. In this chapter, we will consider the Γ-minimax decision rule for selecting populations close to a control and compare it with the Bayes rule and the minimax rule. In so doing, it will be shown how good these rules are.

In Section 1.2, definitions and notations used in this chapter are introduced, and a decision-theoretic formulation of the problem is given. In Section 1.3 and Section 1.4, we derive a Γ-minimax decision rule for both cases when the control parameter θ_0 is known and when it is unknown. It should be pointed out that Randles and Hollander (1971) considered a Γ-minimax procedure for selecting populations better than

a control. When θ_0 is unknown; they applied the Hunt-Stein theorem to prove the Γ-minimaxity of their rules for the component problem. However, the proof given by them does not justify that the Γ-minimax rule of the component problem will give us a Γ-minimax rule for the whole problem. Miesche (1979) gave another technique which can be applied to our problem to solve for the Γ-minimax rule when θ_0 is unknown; this is done in Section 1.4.

In Section 1.5 some optimal properties of Γ-minimax rules are found. In Section 1.6, we generalize the results of Section 1.3 and 1.4, and derive Γ-minimax rules for some more general distributions besides the normal. A Γ-minimax rule for selecting the populations with large entropy is given as an example. In Section 1.7, under the assumption that the prior distributions are $N(\alpha_i, \beta_i^2)$, we find the Bayes rules and we also find the minimax rules. These rules and the Γ-minimax rules found in Section 1.4 are compared in Section 1.8 in terms of the Bayes risk, the maximum risk over Γ, and the overall maximum risk for all possible choices of the prior distributions.

Numerical tables are given for selected values of variables for comparison of these rules. Finally, in Section 1.9, we give an example in which we apply the optimal selection rules. Conclusion about the robustness of each rule discussed in this chapter are also given in Section 1.9.

1.2 Notation and formulation of the problem

Let $\Pi_0, \Pi_1, \ldots, \Pi_k$ be $(k+1)$ independent normal populations with means $\theta_0, \theta_1, \ldots, \theta_k$ and common known variance σ^2, respectively.

Π_0 is the control population, the other populations are defined as good or bad by

<u>Definition 1.2.1.</u> Let $\Delta > 0$, $\varepsilon > 0$ be two given numbers, then

(i) Population Π_i is good iff $|\theta_0-\theta_i| \le \Delta$

(ii) Population Π_i is bad iff $|\theta_0-\theta_i| \ge \Delta + \varepsilon$.

Note that we do not define Π_i as good or bad if $\Delta < |\theta_i-\theta_0| < \Delta + \varepsilon$, which allows us to regard it as an indifference zone between good and bad populations. Throughout this chapter, Δ and ε will be assumed given and fixed. We are interested in selecting as many as possible good populations, and rejecting as many as possible the bad ones. We formulate this problem in the framework of multiple decision theory. Let

$$\Theta = \text{parameter space} = \begin{cases} \{\underset{\sim}{\theta} = (\theta_0,\theta_1,\ldots,\theta_k) \mid -\infty < \theta_i < \infty \text{ for all } i = 0,\ldots,k\} \\ \qquad \text{if } \theta_0 \text{ is unknown} \\ \{\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_k) \mid -\infty < \theta_i < \infty \text{ for all } i = 1,\ldots,k\} \\ \qquad \text{if } \theta_0 \text{ is known} \end{cases}$$

Let $\Theta_G(i) = \{\underset{\sim}{\theta} \in \Theta \mid |\theta_i-\theta_0| \le \Delta\}$, $\Theta_B(i) = \{\underset{\sim}{\theta} \in \Theta \mid |\theta_i-\theta_0| \ge \Delta + \varepsilon\}$. Let $X_{i1}, X_{i2},\ldots, X_{in}$ be the observations from Π_i $(0 \le i \le k)$. Since $\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$ is the sufficient statistic for θ_i and $\{\bar{X}_i\}_{i=0}^{k}$ are independently normally distributed with means $\theta_0, \theta_1,\ldots, \theta_k$ and common known variance $(= \frac{\sigma^2}{n})$, so without loss of generality (wlog)

we can assume that there is only one observation X_i from each population Π_i. Then

$$X_i \sim \frac{1}{\sigma}\phi\left(\frac{x-\theta_i}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2\sigma^2}(x-\theta_i)^2}$$

and

$$\underset{\sim}{X} = \begin{cases} (X_0, X_1, \ldots, X_k) & \text{for } \theta_0 \text{ unknown} \\ (X_1, \ldots, X_k) & \text{for } \theta_0 \text{ known.} \end{cases}$$

The sample space X is defined as follows:

$$X = \begin{cases} \{\underset{\sim}{x} = (x_0, x_1, \ldots, x_k) \mid -\infty < x_i < \infty \quad \text{for all } i=0,1,\ldots k\} \\ \qquad \text{if } \theta_0 \text{ is unknown} \\ \{\underset{\sim}{x} = (x_1, \ldots, x_k) \mid -\infty < x_i < \infty \quad \text{for all } i=1,\ldots,k\} \\ \qquad \text{if } \theta_0 \text{ is known.} \end{cases}$$

Let $(X, \mathcal{B})$ be the usual Lebesque-measurable space. Let

$$D = \{\underset{\sim}{\delta} = (\delta_1, \ldots, \delta_k) \mid \delta_i : X \to [0,1] \text{ is a measurable function, for all } 1 \le i \le k\}.$$

Then D is the set of all selection rules and $\delta_i(\underset{\sim}{x})$ is the probability of selecting Π_i when we observe $\underset{\sim}{X} = \underset{\sim}{x}$.

Let L_1 denote the loss incurred when we fail to select a good population and L_2 the loss for each bad population selected. We define the loss $L(\underset{\sim}{\theta}, \underset{\sim}{\delta})$ of using selection rule $\underset{\sim}{\delta}$ when $\underset{\sim}{\theta}$ is the true state of parameter as follows:

Definition 1.2.2. $L(\underset{\sim}{\theta},\underset{\sim}{\delta}(\underset{\sim}{x})) = \sum_{i=1}^{k} L^{(i)}(\underset{\sim}{\theta},\delta_i(\underset{\sim}{x}))$

$$\text{where } L^{(i)}(\underset{\sim}{\theta},\delta_i(\underset{\sim}{x})) = \begin{cases} L_1(1-\delta_i(\underset{\sim}{x})) & \text{if } \underset{\sim}{\theta} \in \Theta_G(i) \\ L_2\, \delta_i(\underset{\sim}{x}) & \text{if } \underset{\sim}{\theta} \in \Theta_B(i) \\ 0 & \text{otherwise.} \end{cases} \qquad (1.2.1)$$

Finally, we will assume that our partial information is that Π_i has probability λ_i to be good and probability λ_i' to be bad. Also, λ_i and λ_i' are known to us with $0 \le \lambda_i, \lambda_i'$ & $\lambda_i + \lambda_i' \le 1$.

Definition 1.2.3. $\Gamma = \{\tau \mid \tau$ is a prior distribution on Θ &

$$P_\tau[\Theta_G(i)] = \lambda_i,\ P_\tau[\Theta_B(i)] = \lambda_i', \quad \text{for all } 0 \le i \le k\} \qquad (1.2.2)$$

One can see that Γ is the class of all possible prior distributions on Θ which summarizes our information about $\theta_0,\theta_1,\ldots,\theta_k$. Let P_τ denote the Lebesque-Stieljes measure corresponding to τ, then for any Lebesque-measurable set $A \subseteq \Theta$, $P_\tau[A] = \int_A d\tau(\underset{\sim}{\theta})$.

Definition 1.2.4. For all $\tau \in \Gamma$ and $\underset{\sim}{\delta} \in D$, we define

$$r(\tau,\underset{\sim}{\delta}) = E_\tau[R(\underset{\sim}{\theta},\underset{\sim}{\delta})]$$

where $\underset{\sim}{\Theta}$ is a random variable distributed as $\tau(\underset{\sim}{\theta})$ and $R(\underset{\sim}{\theta},\underset{\sim}{\delta}) = E_{\underset{\sim}{\theta}}[L(\underset{\sim}{\theta},\underset{\sim}{\delta}(\underset{\sim}{X}))]$.

A rule $\underset{\sim}{\delta}^* \in D$ is said to be a Γ-minimax rule iff

$$\sup_{\tau\in\Gamma} r(\tau,\underset{\sim}{\delta}^*) = \inf_{\underset{\sim}{\delta}\in D} \sup_{\tau\in\Gamma} r(\tau,\underset{\sim}{\delta}).$$

<u>Definition 1.2.5.</u> For any i $(1 \leq i \leq k)$, the $i^{\underline{th}}$ component problem is to treat the above problem as if we only pay for the loss for wrong decision about Π_i. Hence, the $i^{\underline{th}}$ - component problem is only concerned with $(\Theta, \{\delta_i(\underset{\sim}{x})\}, L^{(i)})$. Similarly, we will use $R^{(i)}(\underset{\sim}{\theta},\delta_i) = E_{\underset{\sim}{\theta}}[L^{(i)}(\underset{\sim}{\theta},\delta_i(\underset{\sim}{X})]$ and $r^{(i)}(\tau,\delta_i) = E_\tau[R^{(i)}(\underset{\sim}{\theta},\delta_i)]$ to denote the risk of $i^{\underline{th}}$ - component problem. We see that $R(\underset{\sim}{\theta},\underset{\sim}{\delta}) = \sum_{i=1}^{k} R^{(i)}(\underset{\sim}{\theta},\delta_i)$ and $r(\tau,\underset{\sim}{\delta}) = \sum_{i=1}^{k} r^{(i)}(\tau,\delta_i)$. This suggests that in order to find the Γ-minimax rule, we may treat the $i^{\underline{th}}$ - component problem separately. In the next section, a Γ-minimax rule is derived for the case when θ_0 is known.

1.3 Derivation of a Γ-minimax rule when θ_0 is known

In this section, θ_0 is treated as known. We consider the $i^{\underline{th}}$ - component problem first.

<u>Lemma 1.3.1.</u> Let $\delta_i(\underset{\sim}{x})$ be an $i^{\underline{th}}$ - component decision rule, if

$$\inf_{\underset{\sim}{\theta}\in\Theta_G(i)} E_{\underset{\sim}{\theta}}[\delta_i(\underset{\sim}{X})] = E_{\theta_i=\theta_0+\Delta}[\delta_i(\underset{\sim}{X})] = E_{\theta_i=\theta_0-\Delta}[\delta_i(\underset{\sim}{X})]$$

and

$$\sup_{\underset{\sim}{\theta}\in\Theta_B(i)} E[\delta_i(\underset{\sim}{X})] = E_{\theta_i=\theta_0+\Delta+\varepsilon}[\delta_i(\underset{\sim}{X})] = E_{\theta_i=\theta_0-\Delta-\varepsilon}[\delta_i(\underset{\sim}{X})], \tag{1.3.1}$$

then for

$$\Gamma_0(i) = \{\tau\in\Gamma \mid P_\tau[\theta_i=\theta_0+\Delta] + P_\tau[\theta_i=\theta_0-\Delta] = \lambda_i$$
$$\text{and} \quad P_\tau[\theta_i=\theta_0+\Delta+\varepsilon] + P_\tau[\theta_i=\theta_0-\Delta-\varepsilon] = \lambda_i^*\},$$

we have

$$\sup_{\tau\in\Gamma} r^{(i)}(\tau,\delta_i) = r^{(i)}(\tau_0,\delta_i) \quad \text{for all} \quad \tau_0 \in \Gamma_0(i).$$

<u>Proof</u>: $\forall\ \tau \in \Gamma$,

$$r^{(i)}(\tau,\delta_i) = \int_{\Theta_G(i)} E_{\underset{\sim}{\theta}}[L_1(1-\delta_i(X)]\ d\tau(\underset{\sim}{\theta})$$

$$+ \int_{\Theta_B(i)} E_{\underset{\sim}{\theta}}[L_2\ \delta_i(\underset{\sim}{X})]\ d\tau(\underset{\sim}{\theta})$$

$$\leq L_1\lambda_i - L_1\lambda_i\ (\inf_{\underset{\sim}{\theta}\in\Theta_G(i)} E_{\underset{\sim}{\theta}}[\delta_i(\underset{\sim}{X})]) + L_2\lambda_i'(\sup_{\underset{\sim}{\theta}\in\Theta_B(i)} E_{\underset{\sim}{\theta}}[\delta_i(\underset{\sim}{X})])$$

$$= L_1\lambda_i - L_1 P_{\tau_0}[\theta_i = \theta_0 + \Delta]\ E_{\theta_i=\theta_0+\Delta}[\delta_i(\underset{\sim}{X})]$$

$$- L_1 P_{\tau_0}[\theta_i=\theta_0-\Delta]\ E_{\theta_i=\theta_0-\Delta}[\delta_i(\underset{\sim}{X})]$$

$$+ L_2 P_{\tau_0}[\theta_i=\theta_0+\Delta+\varepsilon]\ E_{\theta_i=\theta_0+\Delta+\varepsilon}[\delta_i(\underset{\sim}{X})]$$

$$+ L_2 P_{\tau_0}[\theta_i=\theta_0-\Delta-\varepsilon]\ E_{\theta_i=\theta_0-\Delta-\varepsilon}[\delta_i(\underset{\sim}{X})]$$

$$= \int_{\Theta_G(i)} E_{\underset{\sim}{\theta}}[L_1\ (1-\delta_i(\underset{\sim}{X}))]\ d\tau_0(\underset{\sim}{\theta}) + \int_{\Theta_B(i)} E_{\underset{\sim}{\theta}}[L_2\delta_i(\underset{\sim}{X})]\ d\tau_0(\underset{\sim}{\theta})$$

$$= r^{(i)}(\tau_0,\delta_i).$$

The following lemma has been widely used to solve for the Γ-minimax rule. It is stated here without proof.

<u>Lemma 1.3.2.</u> (Randles and Hollander (1971))
If there exists a prior distribution $\tau^* \in \Gamma$ such that the Bayes rule $\delta_i^*(\underset{\sim}{x})$ for the $i\underline{^{th}}$ - component problem wrt τ^* satisfies

$$\sup_{\tau\in\Gamma} r^{(i)}(\tau,\delta_i^*) = r^{(i)}(\tau^*,\delta_i^*) \quad \text{for all } i = 1,2,\ldots,k,$$

then $\underset{\sim}{\delta}^* = (\delta_1^*,\ldots,\delta_k^*)$ is a Γ-minimax decision rule.

Combining Lemma 1.3.1 and Lemma 1.3.2, we get the following theorem:

Theorem 1.3.1. If for $i = 1,2,\ldots,k$, $\delta_i^*(\underset{\sim}{x})$ is a Bayes rule for the $i^{\underline{th}}$ - component problem wrt the same prior distribution $\tau^* \in \bigcap_{i=1}^{k} \Gamma_0(i)$ and assume that δ_i^* satisfies (1.3.1), then $\underset{\sim}{\delta}^* = (\delta_1^*,\ldots,\delta_k^*)$ is a Γ-minimax rule.

Proof: Since $\tau^* \in \Gamma_0(i)$ for all i, hence by Lemma 1.3.1,

$$(1.3.1) \Longrightarrow \sup_{\tau\in\Gamma} r^{(i)}(\tau,\delta_i^*) = r^{(i)}(\tau^*,\delta_i^*) \quad \text{for all } i.$$

Then by Lemma 1.3.2,

$$\delta^* = (\delta_1^*,\ldots,\delta_k^*) \text{ is a } \Gamma\text{-minimax rule.}$$

Remark: For Lemma 1.1.2 and Theorem 1.3.1 to hold, we do not need to assume that the populations are normally distributed. But to satisfy condition (1.3.1), we will restrict ourselves to normal populations from now on. Some results for general distributions will be discussed in Section 1.6.

To verify (1.3.1), some tools which transfer the monotonicity of functions on X to the nonotonicity of function on Θ are needed. We quote some definitions and theorems from Karlin (1968).

Definition 1.3.1. $X \sim f_\theta(x)$ is said to be TP_n (Totally Positive of order n) iff for any $\theta_1 < \theta_2 < \dots < \theta_n$, $x_1 < x_2 < \dots < x_n$, we have

$$K_m\begin{pmatrix} x_1,\dots,x_m \\ \theta_1,\dots,\theta_m \end{pmatrix} = \begin{vmatrix} f_{\theta_1}(x_1) & \dots & f_{\theta_1}(x_m) \\ \vdots & & \vdots \\ f_{\theta_m}(x_1) & \dots & f_{\theta_m}(x_m) \end{vmatrix} \geq 0$$

for all $1 \leq m \leq n$.

Definition 1.3.2. If $X \sim f_\theta(x)$ is TP_n for all $n = 1,2,\dots$, then X is said to be TP (Totally Positive).

Lemma 1.3.3. If $f_\theta(x) = a(\theta)\, b(x)\, e^{\alpha(\theta)\beta(x)}$, where $a(\theta) > 0$, $b(x) > 0$, and $\alpha(\theta)$, $\beta(x)$ are increasing functions, then $f_\theta(x)$ is TP.

Definition 1.3.3. For any real-valued function h, let $S(h)$ denote the number of sign changes of h; we define $S(h) = n$ iff there exist $x_1 < x_2 < \dots < x_{n+1}$ such that either

$$(-1)^{j+1}\, h(x_j) > 0 \;\; \forall \;\; j = 1,2,\dots,n+1$$

or

$$(-1)^{j}\, h(x_j) > 0 \quad \forall \;\; j = 1,2,\dots,n+1,$$

but for any $y_1 < y_2 < \dots < y_{n+1}$, the above two inequalities do not hold.

__Theorem 1.3.2.__ (Karlin) Variation Diminishing Property

If $X \sim f_\theta(x)$ is TP_n and h is a piecewise-continuous function. Let $g(\theta) = E_\theta[h(X)]$, then

$$S(h) \le n - 1 \implies S(g) \le S(h).$$

Furthermore, if $S(g) = S(h) = n - 1$, then g and h change signs in the same order.

__Corollary 1.3.1.__ If $h(x) = I_{[a,b]}(x)$ where I is the indicator function and $X \sim f_\theta(x) = a(\theta)\, b(x)\, e^{\alpha(\theta)\beta(x)}$ with $\alpha(\theta)$, $\beta(x)$ increasing in θ, x, respectively, then if $g(\theta) = E_\theta[h(X)]$ and $g(\theta_0+\theta) = g(\theta_0-\theta)$ for some θ_0, we have g is increasing for $\theta < \theta_0$ and decreasing for $\theta > \theta_0$.

__Proof:__ X TP (by Lemma 1.3.3) $\implies$ X is TP_3.

Now, for $0 < c < 1$, let

$$h_c(x) = h(x) - c \quad\text{and}\quad g_c(\theta) = g(\theta) - c,$$

then

$$E_\theta[h_c(X)] = g_c(\theta).$$

Since $S(h_c) = 2 \le 3 - 1$, we get $S(g_c) \le 2$ (by Theorem 1.3.2) $\forall\, 0 < c < 1$. Now, if g is not increasing for $\theta < \theta_0$, then there exists $\theta_1 < \theta_2 < \theta_0$ and $g(\theta_1) > g(\theta_2)$. Let

$$\theta_1' = 2\theta_0 - \theta_1 \quad\text{and}\quad \theta_2' = 2\theta_0 - \theta_2,$$

we get

$$g(\theta_1') = g(\theta_1) \quad\text{and}\quad g(\theta_2') = g(\theta_2), \quad\text{but}\quad \theta_2' < \theta_1'.$$

Now let

$$c_0 = \frac{1}{2}\,(g(\theta_1) + g(\theta_2)), \quad\text{then}\quad 0 < c_0 < 1.$$

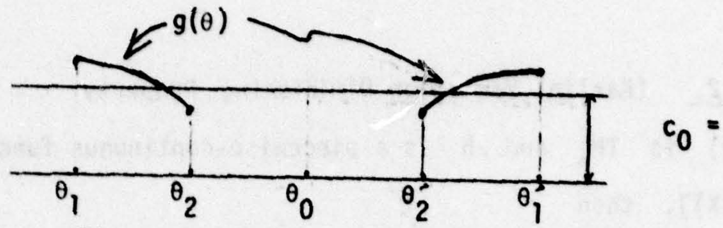


Figure 1. Number of sign changes of function g.

As we can see, $S(g_{c_0}) \geq 2 \Longrightarrow S(g_{c_0}) = 2$. But then by Theorem 1.3.2, g_{c_0} should change sign in the same order as h_c does, which is not true for g_c and h_c. This completes the proof.

Remark: For Corollary 1.3.1 to hold, we only need X to be TP_3. But since we will consider the distributions which are mainly TP, so we have this stronger assumption included.

Corollary 1.3.2. Let $X \sim N(\theta,\sigma^2)$. If

$\delta(x) = I_{[-t+\theta_0,\ t+\theta_0]}(x)$ for some $t > 0$, and $g(\theta) = E_\theta[\delta(X)]$,

then we have

$g(\theta)$ is increasing for $\theta < \theta_0$ and

$g(\theta)$ is decreasing for $\theta > \theta_0$.

Proof: Let $Z \sim N(0,\sigma^2)$, then $Z \sim -Z$.

Now,
$$\begin{aligned} g(\theta+\theta_0) &= Pr[-t + \theta_0 \leq Z + \theta + \theta_0 \leq t + \theta_0] \\ &= Pr[-t - \theta \leq Z \leq t - \theta] \\ &= Pr[-t + \theta \leq -Z \leq t + \theta] \\ &= Pr[-t + \theta \leq Z \leq t + \theta] \\ &= Pr[-t + \theta_0 \leq Z + \theta_0 - \theta \leq t + \theta_0] \\ &= g(\theta_0 - \theta) \end{aligned}$$

then by Corollary 1.3.1, we proved $g(\theta)$ is increasing for $\theta < \theta_0$ and decreasing for $\theta > \theta_0$. This completes the proof.

Note that Corollary 1.3.2. is important for us to justify condition (1.3.1). Now, we turn to the main theorem of this section.

Theorem 1.3.3. Let $X_i \sim N(\theta_i, \sigma^2)$ for $i = 1,2,\ldots,k$ be independent random variable with σ^2 known. If $\underset{\sim}{\delta}^* = (\delta_1^*, \ldots, \delta_k^*)$ where

$$\delta_i^*(x_i) = I_{[-t_i + \theta_0,\ \theta_0 + t_i]}(x_i)$$

and $\pm t_i$, are determined by the equation

$$L_2\lambda_i'[\phi(\frac{t_i+\Delta+\varepsilon}{\sigma}) + \phi(\frac{t_i-\Delta-\varepsilon}{\sigma})]$$

$$= L_1\lambda_i\ [\phi(\frac{t_i+\Delta}{\sigma}) + \phi(\frac{t_i-\Delta}{\sigma})]\ , \tag{1.3.2}$$

then $\underset{\sim}{\delta}^*$ is a Γ-minimax rule.

Proof: We define τ^* to be the prior in Γ such that $\theta_1, \theta_2, \ldots, \theta_k$ are independent and satisfy

$$P_{\tau^*}[\theta_i = \theta_0 - \Delta - \varepsilon] = P_{\tau^*}[\theta_i = \theta_0 + \Delta + \varepsilon] = \frac{\lambda_i'}{2}$$

$$P_{\tau^*}[\theta_i = \theta_0 - \Delta] = P_{\tau^*}[\theta_i = \theta_0 + \Delta] = \frac{\lambda_i}{2}$$

$$P_{\tau^*}[\theta_i = \theta_0 + \Delta + \frac{\varepsilon}{2}] = 1 - \lambda_i - \lambda_i'\ ,$$

for all $i = 1,2,\ldots,k$. Then it is easily seen that $\tau^* \in \bigcap_{i=1}^{k} \Gamma_0(i)$.

Now, let

$$f_{\underset{\sim}{\theta}}(\underset{\sim}{x}) = \prod_{i=1}^{k} f_{\theta_i}(x_i)$$

where

$$f_{\theta_i}(x_i) = \frac{1}{\sigma}\ \phi\left(\frac{x_i - \theta_i}{\sigma}\right),$$

then we have

$$r^{(i)}(\tau^*,\delta_i) = \int_\Theta \int_X L^{(i)}(\underset{\sim}{\theta},\delta_i(\underset{\sim}{x}))\ f_{\underset{\sim}{\theta}}(\underset{\sim}{x})\ d\underset{\sim}{x}\ d\tau^*(\underset{\sim}{\theta})$$

$$= \int_{|\theta_i-\theta_0|\le\Delta} \int_X L_1(1-\delta_i(\underset{\sim}{x}))\ f_{\underset{\sim}{\theta}}(\underset{\sim}{x})\ d\underset{\sim}{x}\ d\tau^*(\underset{\sim}{\theta})$$

$$+ \int_{|\theta_i-\theta_0|\ge\Delta+\varepsilon} \int_X L_2\ \delta_i(\underset{\sim}{x})\ f_{\underset{\sim}{\theta}}(\underset{\sim}{x})\ d\underset{\sim}{x}\ d\tau^*(\underset{\sim}{\theta})$$

$$= \int_X L_1(1-\delta_i(\underset{\sim}{x})) \sum_{\underset{\sim}{\theta}\in\{\theta_i=\theta_0-\Delta\}} f_{\underset{\sim}{\theta}}(\underset{\sim}{x})\ P_{\tau^*}(\underset{\sim}{\theta})\ d\underset{\sim}{x}$$

$$+ \int_X L_1(1-\delta_i(\underset{\sim}{x})) \sum_{\underset{\sim}{\theta}\in\{\theta_i=\theta_0+\Delta\}} f_{\underset{\sim}{\theta}}(\underset{\sim}{x})\ P_{\tau^*}(\underset{\sim}{\theta})\ d\underset{\sim}{x}$$

$$+ \int_X L_2\delta_i(\underset{\sim}{x}) \sum_{\underset{\sim}{\theta}\in\{\theta_i=\theta_0-\Delta-\varepsilon\}} f_{\underset{\sim}{\theta}}(\underset{\sim}{x})\ P_{\tau^*}(\underset{\sim}{\theta})\ d\underset{\sim}{x}$$

$$+ \int_X L_2\delta_i(\underset{\sim}{x}) \sum_{\underset{\sim}{\theta}\in\{\theta_i=\theta_0+\Delta+\varepsilon\}} f_{\underset{\sim}{\theta}}(\underset{\sim}{x})\ P_{\tau^*}(\underset{\sim}{\theta})\ d\underset{\sim}{x}.$$

Now we may notice that

$$\frac{\sum_{\underset{\sim}{\theta}\in\{\theta_i=\theta_0-\Delta\}} f_{\underset{\sim}{\theta}}(\underset{\sim}{x})\ P_{\tau^*}(\underset{\sim}{\theta})}{f_{\theta_0-\Delta}(x_i)\ \frac{\lambda_i}{2}} = \frac{\sum_{\underset{\sim}{\theta}\in\{\theta_i=\theta_0+\Delta\}} f_{\underset{\sim}{\theta}}(\underset{\sim}{x})\ P_{\tau^*}(\underset{\sim}{\theta})}{f_{\theta_0+\Delta}(x_i)\ \frac{\lambda_i}{2}}$$

$$= \frac{\sum_{\underset{\sim}{\theta}\in\{\theta_i=\theta_0-\Delta-\varepsilon\}} f_{\underset{\sim}{\theta}}(\underset{\sim}{x})\ P_{\tau^*}(\underset{\sim}{\theta})}{f_{\theta_0-\Delta-\varepsilon}(x_i)\ \frac{\lambda_i^*}{2}} = \frac{\sum_{\underset{\sim}{\theta}\in\{\theta_i=\theta_0+\Delta+\varepsilon\}} f_{\underset{\sim}{\theta}}(\underset{\sim}{x})\ P_{\tau^*}(\underset{\sim}{\theta})}{f_{\theta_0+\Delta+\varepsilon}(x_i)\ \frac{\lambda_i^*}{2}}.$$

Any of the above four expressions is denoted by $c(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_k)$. Hence,

$$r^{(i)}(\tau^*,\delta_i) = \int_X [L_1(1-\delta_i(\underset{\sim}{x})) \frac{\lambda_i}{2} f_{\theta_0-\Delta}(x_i) +$$

$$L_1(1-\delta_i(\underset{\sim}{x})) \frac{\lambda_i}{2} f_{\theta_0+\Delta}(x_i) + L_2\delta_i(\underset{\sim}{x}) \frac{\lambda_i'}{2} f_{\theta_0-\Delta-\varepsilon}(x_i)$$

$$+ L_2\delta_i(\underset{\sim}{x}) \frac{\lambda_i'}{2} f_{\theta_0+\Delta+\varepsilon}(x_i)]\, c(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_k)\, d\underset{\sim}{x}$$

$$= \int_X L_1 c(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_k)\, [f_{\theta_0-\Delta}(x_i) + f_{\theta_0+\Delta}(x_i)] \frac{\lambda_i}{2}\, d\underset{\sim}{x}$$

$$+ \int_X \left\{ \frac{L_2\lambda_i'}{2} [f_{\theta_0-\Delta-\varepsilon}(x_i) + f_{\theta_0+\Delta+\varepsilon}(x_i)] - \frac{L_1\lambda_i}{2} [f_{\theta_0-\Delta}(x_i)+f_{\theta_0+\Delta}(x_i)] \right\}$$

$$\cdot\, \delta_i(\underset{\sim}{x}) \cdot c(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_k)\, d\underset{\sim}{x}\ .$$

Thus, we find that the Bayes rule is given by

$$\delta_i^*(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } L_2\lambda_i'[f_{\theta_0-\Delta-\varepsilon}(x_i) + f_{\theta_0+\Delta+\varepsilon}(x_i)] \le L_1\lambda_i[f_{\theta_0-\Delta}(x_i)+f_{\theta_0+\Delta}(x_i)] \\ 0 & \text{if } L_2\lambda_i'[f_{\theta_0-\Delta-\varepsilon}(x_i)+f_{\theta_0+\Delta+\varepsilon}(x_i)] > L_1\lambda_i[f_{\theta_0-\Delta}(x_i)+f_{\theta_0+\Delta}(x_i)]. \end{cases} \tag{1.3.3}$$

Let

$$h_i(x_i) = \frac{L_2\lambda_i' \,[f_{\theta_0-\Delta-\varepsilon}(x_i) + f_{\theta_0+\Delta+\varepsilon}(x_i)]}{L_1\lambda_i\,[f_{\theta_0-\Delta}(x_i) + f_{\theta_0+\Delta}(x_i)]}$$

$$= \frac{L_2\lambda_i'}{L_1\lambda_i}\ \frac{e^{-\frac{1}{2\sigma^2}(\Delta+\varepsilon)^2}\left[e^{-\frac{1}{\sigma^2}(x_i-\theta_0)(\Delta+\varepsilon)} + e^{\frac{1}{\sigma^2}(x_i-\theta_0)(\Delta+\varepsilon)}\right]}{e^{-\frac{1}{2\sigma^2}\Delta^2}\left[e^{-\frac{1}{\sigma^2}(x_i-\theta_0)\Delta} + e^{\frac{1}{\sigma^2}(x_i-\theta_0)\Delta}\right]}$$

21

$$= \frac{L_2\lambda_i'}{L_1\lambda_i} e^{-\frac{1}{2\sigma^2}(2\Delta+\varepsilon)\varepsilon} \frac{\cosh\left[\frac{1}{\sigma^2}(x_i-\theta_0)(\Delta+\varepsilon)\right]}{\cosh\left[\frac{1}{\sigma^2}(x_i-\theta_0)\Delta\right]} .$$

Since $\cosh(y) = \cosh(-y)$, we see that

$$h_i(\theta_0+x_i) = h_i(\theta_0-x_i). \tag{1.3.4}$$

Moreover, $\frac{d}{dx}\frac{\cosh ax}{\cosh bx} = \frac{b}{\cosh^2 bx}\sinh[(a-b)x]$.

Hence, for $a > b > 0$, $\frac{\cosh ax}{\cosh bx}$ is decreasing for $x < 0$ and increasing if $x > 0$. This shows that $h_i(x_i)$ is decreasing for $x_i < \theta_0$ and increasing for $x_i > \theta_0$. Hence it follows that $h_i(x_i) = 1$ has two solutions, $\pm t_i + \theta_0$. Then, from (1.3.3),

$$\delta_i^*(x) = \delta_i^*(x_i) = \begin{cases} 1 & \text{if } h(x_i) \le 1 \\ 0 & \text{if } h(x_i) > 1 \end{cases}$$

$$= \begin{cases} 1 & \text{if } -t_i + \theta_0 \le x_i \le \theta_0 + t_i \\ 0 & \text{otherwise} \end{cases} \quad \text{by (1.3.4)}$$

and t_i's satisfy (1.3.2).

Finally, let $g(\theta_i) = E_{\theta_i}[\delta_i^*(X_i)]$, then from Corollary 1.3.2, $g(\theta_i+\theta_0) = g(\theta_0-\theta_i)$ and g is increasing for $\theta_i < \theta_0$ and decreasing for $\theta_i > \theta_0$. This proves that

$$\inf_{\underset{\sim}{\theta} \in \Theta_G(i)} g(\theta_i) = g(\theta_0+\Delta) = g(\theta_0-\Delta)$$

and

$$\sup_{\underset{\sim}{\theta} \in \Theta_B(i)} g(\theta_i) = g(\theta_0+\Delta+\varepsilon) = g(\theta_0-\Delta-\varepsilon),$$

i.e., condition (1.3.1) is satisfied. Now, Theorem 1.3.1. shows that $\underset{\sim}{\delta}^* = (\delta_1^*,\ldots,\delta_k^*)$ is a Γ-minimax decision rule. This completes the proof.

Note that it may happen (1.3.2) does not have a solution for some λ_i, λ_i', Δ, ε. In this case, the Γ-minimax rule implies that all populations are bad.

The solutions t_i depend on λ_i and λ_i' only through their ratio $\nu_i = \lambda_i/\lambda_i'$ (see (1.3.2)), hence $\underset{\sim}{\delta}^*$ is actually a Γ-minimax rule for $\Gamma = \{\tau \mid P_\tau[\Theta_G(i)] / P_\tau[\Theta_B(i)] = \nu_i,$ for $i = 1,2,\ldots,k\}$.

1.4 Derivation of a restricted Γ-minimax rule when θ_0 is unknown

In this section, θ_0 will be treated as an unknown parameter. As mentioned in Section 1.2, Θ, X, and $\underset{\sim}{X}$ will include one more component and one observation X_0 is taken from Π_0.

<u>Definition 1.4.1.</u> Let $D_1 \subseteq D$ be the class of rules such that the $i^{\underline{th}}$ decision rule depends only on X_0 and X_i, i.e.,

$$D_1 = \{\underset{\sim}{\delta} = (\delta_1,\ldots,\delta_k) \in D \mid \delta_i = \delta_i(x_0,x_i) \;\; \forall\; 1 \le i \le k\}. \tag{1.4.1}$$

We derive a Γ-minimax rule in the class D_1. The problem whether our rule is also Γ-minimax in D when θ_0 is unknown is not solved.

Ferguson (1967, P. 90) gives two theorems to provide solutions for the minimax rule. Lemma 1.3.2. is similar to his Theorem 1 to solve for a Γ-minimax rule, and the following lemma (due to Miescke) is similar to Ferguson's Theorem 2.

Lemma 1.4.1. (Miescke) If $\{\tau_n \in \Gamma\}_{n=1}^{\infty}$ is a sequence of priors and $\{\delta_{in}^0\}_{n=1}^{\infty}$ is a sequence of Bayes rules corresponding to τ_n for the $i^{\underline{th}}$ - component problem for all $i = 1,2,\ldots,k$, then $\underset{\sim}{\delta}^0 = (\delta_1^0,\ldots,\delta_k^0)$ is a Γ-minimax rule iff

$$\lim_{n\to\infty} \inf r^{(i)}(\tau_n,\delta_{in}^0) \ge \sup_{\tau\in\Gamma} r^{(i)}(\tau,\delta_i^0) \quad \text{for all } i = 1,\ldots,k. \tag{1.4.2}$$

Proof: Let $\underset{\sim}{\delta} = (\delta_1,\ldots,\delta_k)$ be any selection rule. Then

$$\sup_{\tau\in\Gamma} r(\tau,\underset{\sim}{\delta}) = \sup_{\tau\in\Gamma} \sum_{i=1}^{k} r^{(i)}(\tau,\delta_i)$$

$$\ge \sup_{n\in\mathbb{N}} \sum_{i=1}^{k} r^{(i)}(\tau_n,\delta_i) \ge \sup_{n\in\mathbb{N}} \sum_{i=1}^{k} r^{(i)}(\tau_n,\delta_{in}^0)$$

$$\ge \lim_{n\to\infty} \inf \sum_{i=1}^{k} r^{(i)}(\tau_n,\delta_{in}^0) \ge \sum_{i=1}^{k} \lim_{n\to\infty} \inf r^{(i)}(\tau_n,\delta_{in}^0)$$

$$\ge \sup_{\tau\in\Gamma} \sum_{i=1}^{k} r^{(i)}(\tau,\delta_i^0) = \sup_{\tau\in\Gamma} r(\tau,\underset{\sim}{\delta}^0)\ .$$

This proves $\underset{\sim}{\delta}^0$ is a Γ-minimax rule.

To use the preceeding lemma, we need to find a sequence of priors and their rules so that (1.4.2) holds. Now, each prior distribution τ on $\Theta_0 \times \Theta_1 \times \ldots \times \Theta_k$ can be specified by the marginal distribution T_0 on Θ_0 and the conditional distribution ω_{θ_0} on $\Theta_1 \times \ldots \times \Theta_k$ given $\Theta_0 = \theta_0$. We will use $\tau = (T_0,\omega_{\theta_0})$ to denote such prior distributions. Let

$$\tau_n = (T_n,\omega_{\theta_0})$$

where

(i) T_n is the marginal distribution on Θ_0 which is assumed to be uniform over $[-n,n]$.

(ii) Given $\Theta_0 = \theta_0$, i.e., under ω_{θ_0}, $\theta_1,\ldots,\theta_k$ are independent and

$$P_{\omega_{\theta_0}}[\theta_i = \theta_0 - \Delta - \varepsilon \mid \theta_0] = P_{\omega_{\theta_0}}[\theta_i = \theta_0 + \Delta + \varepsilon \mid \theta_0] = \frac{\lambda_i'}{2}$$

$$P_{\omega_{\theta_0}}[\theta_i = \theta_0 - \Delta \mid \theta_0] = P_{\omega_{\theta_0}}[\theta_i = \theta_0 + \Delta \mid \theta_0] = \frac{\lambda_i}{2}$$

$$P_{\omega_{\theta_0}}[\theta_i = \theta_0 + \Delta + \frac{\varepsilon}{2} \mid \theta_0] = 1 - \lambda_i - \lambda_i'.$$

We will also use the notation $\omega^i_{\theta_0}$ to denote the conditional marginal distribution of θ_i under ω_{θ_0}.

Lemma 1.4.2. Let τ_n be defined as above, then under the loss function as defined by (1.2.1), the Bayes rule in the class D_1 wrt τ_n for the $i^{\underline{th}}$ - component problem is

$$\delta^0_{in}(x_0,x_i) = \begin{cases} 1 & \text{iff } \lambda_i' L_2 \int_{-n}^{n} [f_{\theta_0+\Delta+\varepsilon}(x_i) + f_{\theta_0-\Delta-\varepsilon}(x_i)] f_{\theta_0}(x_0) d\theta_0 \\ & \leq \lambda_i L_1 \int_{-n}^{n} [f_{\theta_0+\Delta}(x_i) + f_{\theta_0-\Delta}(x_i)] f_{\theta_0}(x_0) d\theta_0 \\ 0 & \quad > \end{cases} \tag{1.4.3}$$

where $f_\theta(x) = \frac{1}{\sigma}\,\phi(\frac{x-\theta}{\sigma})$.

<u>Proof</u>: $r^{(i)}(\tau_n,\delta_i) = \int_{-n}^{n}\Big\{\int_{|\theta_i-\theta_0|\le\Delta} L_1(1-E_{(\theta_0,\theta_i)}[\delta_i(X_0,X_i)])d\omega_{\theta_0}^{i}(\theta_i)$

$+\int_{|\theta_i-\theta_0|\ge\Delta+\varepsilon} L_2E_{(\theta_0,\theta_i)}[\delta_i(X_0,X_i)]d\omega_{\theta_0}^{i}(\theta_i)\Big\}\, dT_n(\theta_0)$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-n}^{n}\Big\{\frac{\lambda_i}{2}L_1(1-\delta_i(x_0,x_i))[f_{\theta_0+\Delta}(x_i) + f_{\theta_0-\Delta}(x_i)]f_{\theta_0}(x_0)$$

$$+ \frac{\lambda_i'}{2} L_2\delta_i(x_0,x_i)[f_{\theta_0+\Delta+\varepsilon}(x_i) + f_{\theta_0-\Delta-\varepsilon}(x_i)]\Big\}\, dT_n(\theta_0)dx_idx_0$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-n}^{n}\frac{\lambda_i}{2} L_1[f_{\theta_0+\Delta}(x_i) + f_{\theta_0-\Delta}(x_i)]f_{\theta_0}(x_0)\, dT_n(\theta_0)dx_idx_0$$

$$+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta_i(x_0,x_i)\Big\{\frac{\lambda_i'L_2}{2}\int_{-n}^{n}[f_{\theta_0+\Delta+\varepsilon}(x_i) + f_{\theta_0-\Delta-\varepsilon}(x_i)]f_{\theta_0}(x_0)\frac{1}{2n}d\theta_0$$

$$- \frac{\lambda_iL_1}{2}\int_{-n}^{n}[f_{\theta_0+\Delta}(x_i) + f_{\theta_0-\Delta}(x_i)]f_{\theta_0}(x_0)\frac{1}{2n}d\theta_0\Big\}\, dx_idx_0.$$

Hence, the Bayes rule is as shown in (1.4.3).

Now we find the Bayes risk of δ_{in}^0 wrt τ_n is

$$r^{(i)}(\tau_n,\delta_{in}^0)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Big\{[1-\delta_{in}^0(x_0,x_i)]\frac{\lambda_iL_1}{2}\int_{-n}^{n}[f_{\theta_0+\Delta}(x_i)+f_{\theta_0-\Delta}(x_i)]f_{\theta_0}(x_0)dT_n(\theta_0)$$

$$+ \delta_{in}^0(x_0,x_i)\,\frac{\lambda_i'L_2}{2}\int_{-n}^{n}[f_{\theta_0+\Delta+\varepsilon}(x_i)+f_{\theta_0-\Delta-\varepsilon}(x_i)]f_{\theta_0}(x_0)dT_n(\theta_0)\Big\}dx_idx_0$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\min\Big\{\frac{\lambda_iL_1}{2}\int_{-n}^{n}[f_{\theta_0+\Delta}(x_i)+f_{\theta_0-\Delta}(x_i)]f_{\theta_0}(x_0)dT_n(\theta_0)\ ,$$

$$\frac{\lambda_i'L_2}{2}\int_{-n}^{n}[f_{\theta_0+\Delta+\varepsilon}(x_i)+f_{\theta_0-\Delta-\varepsilon}(x_i)]f_{\theta_0}(x_0)dT_n(\theta_0)\Big\}\, dx_idx_0\ .$$

If we consider the change of variables $\begin{cases} x_i = ny_i + y_0 \\ x_0 = ny_i - y_0 \end{cases}$

of the two outside integrals, then change $\theta_0 = ny_i - \eta_0$ for the inside integral. Since $\left| \frac{\partial(x_i,x_0)}{\partial(y_i,y_0)} \right| = 2n$ and $f_\theta(x) = f_0(x-\theta)$, we get

$$r^{(i)}(\tau_n,\delta^0_{in}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \min\{\frac{\lambda_i L_1}{2}\int_{n(y_i-1)}^{n(y_i+1)} [f_0(y_0+\eta_0-\Delta)+f_0(y_0+\eta_0+\Delta)]$$

$$f_0(\eta_0-y_0)\,d\eta_0,$$

$$\frac{\lambda_i' L_2}{2}\int_{n(y_i-1)}^{n(y_i+1)} [f_0(y_0+\eta_0-\Delta-\varepsilon)+f_0(y_0+\eta_0+\Delta+\varepsilon)]f_0(\eta_0-y_0)d\eta_0\}dy_i dy_0$$

$$\geq \int_{-\infty}^{\infty}\int_{-1}^{1}\min\{\frac{\lambda_i L_1}{2}\int_{n(y_i-1)}^{n(y_i+1)} [f_0(y_0+\eta_0-\Delta)+f_0(y_0+\eta_0+\Delta)]f_0(\eta_0-y_0)d\eta_0,$$

$$\frac{\lambda_i' L_2}{2}\int_{n(y_i-1)}^{n(y_i+1)} [f_0(y_0+\eta_0-\Delta-\varepsilon)+f_0(y_0+\eta_0+\Delta+\varepsilon)]f_0(\eta_0-y_0)d\eta_0\}dy_i dy_0. \tag{1.4.4}$$

To find a lower bound of $\liminf_{n\to\infty} r(\tau_n,\delta^0_{in})$, we need to use the following facts:

(i) Fatou's Lemma.

(ii) $\frac{1}{\sqrt{2}} f_0(\frac{z-b}{\sqrt{2}}) = \int_{-\infty}^{\infty} f_0(x-z)f_0(x-b)dx$, where $f_0(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$

From (i), (ii) and (1.4.4), we get

$$\liminf_{n\to\infty} r^{(i)}(\tau_n,\delta^0_{in}) \geq \int_{-\infty}^{\infty}\int_{-1}^{1}\min\{\frac{\lambda_i L_1}{2\sqrt{2}}[f_0(\frac{2y_0-\Delta}{\sqrt{2}})+f_0(\frac{2y_0+\Delta}{\sqrt{2}})],$$

$$\frac{\lambda_i' L_2}{2\sqrt{2}} [f_0(\frac{2y_0-\Delta-\varepsilon}{\sqrt{2}}) + f_0(\frac{2y_0+\Delta+\varepsilon}{2})]\}dy_i dy_0$$

$$= \int_{-\infty}^{\infty} \min\{\frac{\lambda_i L_1}{2\sqrt{2}} [f_0(\frac{2y_0-\Delta}{\sqrt{2}}) + f_0(\frac{2y_0+\Delta}{\sqrt{2}})], \frac{\lambda_i' L_2}{2\sqrt{2}} [f_0(\frac{2y_0-\Delta-\varepsilon}{\sqrt{2}} + f_0(\frac{2y_0+\Delta+\varepsilon}{\sqrt{2}})]\} d(2y_0)$$

$$= \int_{-\infty}^{\infty} \min\{\frac{\lambda_i L_1}{2\sqrt{2}} [f_0(\frac{y_0-\Delta}{\sqrt{2}})+f_0(\frac{y_0+\Delta}{\sqrt{2}})], \frac{\lambda_i' L_2}{2\sqrt{2}} [f_0(\frac{y_0-\Delta-\varepsilon}{\sqrt{2}})+f_0(\frac{y_0+\Delta+\varepsilon}{\sqrt{2}})]\}dy_0. \tag{1.4.5}$$

Now, we are ready to prove the following theorem.

Theorem 1.4.1. If θ_0 is unknown, let $L(\underset{\sim}{\theta},\underset{\sim}{\delta}(\underset{\sim}{x}))$, Γ and D_1 be as defined in (1.2.1), (1.2.2) and (1.4.1), respectively, then the Γ-minimax rule in D_1 is given by

$$\underset{\sim}{\delta}^0 = (\delta_1^0,\ldots,\delta_k^0),$$

where

$$\delta_i^0(x_i-x_0) = \begin{cases} 1 & \text{if } L_2\lambda_i'[\phi(\frac{(x_i-x_0)+\Delta+\varepsilon}{\sqrt{2}\,\sigma})+ \phi(\frac{(x_i-x_0)-\Delta-\varepsilon}{\sqrt{2}\,\sigma})] \\ & \quad \le L_1\lambda_i[\phi(\frac{(x_i-x_0)+\Delta}{\sqrt{2}\,\sigma}) + \phi(\frac{(x_i-x_0)-\Delta}{\sqrt{2}\,\sigma})] \\ 0 & \quad > \end{cases} \tag{1.4.6}$$

Proof: Let $Y_i = X_i-X_0$, then $Y_i \sim N(\eta_i,\sigma'^2)$, where $\eta_i = \theta_i-\theta_0$ and $\sigma' = \sqrt{2}\sigma$. Let $g(\eta_i) = E_{\eta_i}[\delta_i^0(Y_i)]$, then as was shown in the proof of Theorem 1.3.3., $g(\eta_i)$ is increasing for $\eta_i < 0$ and decreasing for $\eta_i > 0$, and $g(\eta_i) = g(-\eta_i)$, so

$$\sup_{|\eta_i| \geq \Delta+\varepsilon} g(\eta_i) = g(\Delta+\varepsilon) = g(-\Delta-\varepsilon)$$

and

$$\inf_{|\eta_i| \leq \Delta} g(\eta_i) = g(\Delta) = g(-\Delta).$$

Now, $\forall \tau = (T_0, \omega_{\theta_0}) \in \Gamma$,

$$r^{(i)}(\tau,\delta_i^0) = \int_{-\infty}^{\infty} \{\int_{|\theta_i-\theta_0| \leq \Delta} L_1(1-E_{\theta_i-\theta_0}[\delta_i^0(x_i-x_0)])d\omega_{\theta_0}^i(\theta_i)$$

$$+ \int_{|\theta_i-\theta_0| \geq \Delta+\varepsilon} L_2 E_{\theta_i-\theta_0}[\delta_i^0(x_i-x_0)]d\omega_{\theta_0}^i(\theta_i)\}d\tau_0(\theta_0)$$

$$\leq L_1(1-\inf_{|\theta_i-\theta_0| \leq \Delta} E_{\theta_i-\theta_0}[\delta_i^0(x_i-x_0)])\int_{-\infty}^{\infty}\int_{|\theta_i-\theta_0| \leq \Delta} d\omega_{\theta_0}^i(\theta_i)d\tau_0(\theta_0)$$

$$+ L_2 \sup_{|\theta_i-\theta_0| \geq \Delta+\varepsilon} E_{\theta_i-\theta_0}[\delta_i^0(x_i-x_0)]\int_{-\infty}^{\infty}\int_{|\theta_i-\theta_0| \geq \Delta+\varepsilon} d\omega_{\theta_0}^i(\theta_i)d\tau_0(\theta_0)$$

$$= L_2\lambda_i(1-\inf_{|\eta_i| \leq \Delta} g(\eta_i))+L_2\lambda_i' \sup_{|\eta_i| \geq \Delta+\varepsilon} g(\eta_i)$$

$$= L_1\lambda_i(1-\frac{g(\Delta)+g(-\Delta)}{2}) + L_2\lambda_i' \frac{g(\Delta+\varepsilon)+g(-\Delta-\varepsilon)}{2}$$

$$= \frac{L_1\lambda_i}{2}(1-E_\Delta[\delta_i^0(Y_i)]+1-E_{-\Delta}[\delta_i^0(Y_i)])+ \frac{L_2\lambda_i'}{2}(E_{\Delta+\varepsilon}[\delta_i^0(Y_i)]+E_{-\Delta-\varepsilon}[\delta_i^0(Y_i)])$$

$$= \int_{-\infty}^{\infty} \frac{L_1\lambda_i}{2}(1-\delta_i^0(y_i))[\phi(\frac{y_i-\Delta}{\sigma'})+\phi(\frac{y_i+\Delta}{\sigma'})] \frac{1}{\sigma'}$$

$$+ \frac{L_2\lambda_i'}{2} \delta_i^0(y_i)[\phi(\frac{y_i-\Delta-\varepsilon}{\sigma'})+\phi(\frac{y_i+\Delta+\varepsilon}{\sigma'})] \frac{1}{\sigma'} dy_i$$

$$= \int_{-\infty}^{\infty} \min\{\frac{L_1\lambda_i}{2}[\phi(\frac{y_i-\Delta}{\sigma'})+\phi(\frac{y_i+\Delta}{\sigma'})] \frac{1}{\sigma'}, \frac{L_2\lambda_i'}{2}[\phi(\frac{y_i-\Delta-\varepsilon}{\sigma'})+\phi(\frac{y_i+\Delta+\varepsilon}{\sigma'})] \frac{1}{\sigma'}\}dy_i$$

$$= \int_{-\infty}^{\infty} \min\{\frac{L_1\lambda_i}{2\sqrt{2}}[f_0(\frac{y_i-\Delta}{\sqrt{2}})+f_0(\frac{y_i+\Delta}{\sqrt{2}})],\ \frac{L_2\lambda_i'}{2\sqrt{2}}[f_0(\frac{y_i-\Delta-\varepsilon}{\sqrt{2}})+f_0(\frac{y_i+\Delta+\varepsilon}{\sqrt{2}})]\}dy_i$$

$$\leq \liminf_{n\to\infty} r(\tau_n,\delta_{in}^0) \quad \text{by (1.4.5).}$$

Hence, we have proved that $\sup_{\tau\in\Gamma} r^{(i)}(\tau,\delta_i^0) \leq \liminf_{n\to\infty} r(\tau_n,\delta_{in}^0)$ for all $i = 1,2,\ldots,k$. By Lemma 1.4.1., we conclude $\underset{\sim}{\delta}^0 = (\delta_1^0,\ldots,\delta_k^0)$ is a Γ-minimax rule in D_1. This completes the proof.

By the same reason as mentioned in the last paragraph of Section 1.3, $\underset{\sim}{\delta}^0$ is also a Γ-minimax rule in D_1 for

$$\Gamma = \{\tau|P_\tau[\Theta_G(i)]/P_\tau[\Theta_B(i)] = \nu_i, \text{ for all } i = 0,1,\ldots,k\},$$

where $\nu_i = \lambda_i/\lambda_i'$ for $i = 0,1,\ldots,k$.

1.5 Optimal properties of the Γ-minimax rule

As mentioned in Section 1.2, wlog we can reduce the sample size to 1 for each population. If, in fact, we observe $X_{i1},\ldots,X_{in}$ from π_i, the Γ-minimax rule remains the same with the substitution of X_i by $\bar{X}_{in}$ ($\bar{X}_{in} = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$). We now prove the following theorem.

Theorem 1.5.1. When θ_0 is known,

$$\lim_{n\to\infty} \inf_{|\theta_i-\theta_0| \leq \Delta} E_{\underset{\sim}{\theta}}[\delta_i^*(\bar{X}_{in})] = 1$$

and

$$\lim_{n\to\infty} \sup_{|\theta_i-\theta_0| \geq \Delta+\varepsilon} E_{\underset{\sim}{\theta}}[\delta_i^*(\bar{X}_{in})] = 0.$$

The above theorem says that as n becomes sufficiently large, the

probability of selecting a good population approaches 1 uniformly, and the probability of selecting a bad population goes to 0 uniformly.

To prove Theorem 1.5.1, we need the following lemma.

<u>Lemma 1.5.1</u>. For any sample size n, the Γ-minimax rule $\underset{\sim}{\delta}^* = (\delta_1^*,\ldots,\delta_k^*)$ can be written as

$$\delta_i^*(\bar{x}_{in}) = I_{[-t_i(n),t_i(n)]}(\bar{x}_{in}-\theta_0),$$

then

$$\lim_{n\to\infty} t_i(n) = \Delta + \frac{\varepsilon}{2}\,.$$

<u>Proof</u>: From the proof of Theorem 1.3.3., we know $t_i(n)$ is the positive root of the equation

$$h_n(x) = \frac{L_2\lambda_i'}{L_1\lambda_i} e^{-\frac{n}{2\sigma^2}(2\Delta+\varepsilon)\varepsilon} \frac{\cosh[\frac{n}{\sigma^2}(\Delta+\varepsilon)x]}{\cosh[\frac{n}{\sigma^2}\Delta x]} - 1 = 0.$$

Now, consider

$$f_n(x) = \frac{L_2\lambda_i'}{L_1\lambda_i} e^{-\frac{n}{2\sigma^2}(2\Delta+\varepsilon)\varepsilon} \frac{e^{\frac{n}{\sigma^2}(\Delta+\varepsilon)x}}{e^{\frac{n}{\sigma^2}\Delta x}} - 1$$

$$g_n(x) = \frac{L_2\lambda_i'}{L_1\lambda_i} e^{-\frac{n}{2\sigma^2}(2\Delta+\varepsilon)\varepsilon} \frac{e^{\frac{n}{\sigma^2}(\Delta+\varepsilon)x}}{2e^{\frac{n}{\sigma^2}\Delta x}} - 1.$$

Because for x > 0, we have

$$\frac{e^{\frac{n}{\sigma^2}(\Delta+\varepsilon)x}}{2e^{\frac{n}{\sigma^2}\Delta x}} \le \frac{e^{\frac{n}{\sigma^2}(\Delta+\varepsilon)x}+e^{-\frac{n}{\sigma^2}(\Delta+\varepsilon)x}}{e^{\frac{n}{\sigma^2}\Delta x}+e^{-\frac{n}{\sigma^2}\Delta x}} \le \frac{e^{\frac{n}{\sigma^2}(\Delta+\varepsilon)x}}{e^{\frac{n}{\sigma^2}\Delta x}},$$

so $g_n(x) \le h_n(x) \le f_n(x)$ for $x > 0$. Let $r_i(n)$ and $s_i(n)$ be the only positive roots of $g_n(x) = 0$ and $f_n(x) = 0$, respectively, we get

$$r_i(n) \ge t_i(n) \ge s_i(n),$$

but

$$r_i(n) = \frac{\frac{n}{2\sigma^2}(2\Delta+\varepsilon)\varepsilon - \ln\left(\frac{L_2\lambda_i'}{2L_1\lambda_i}\right)}{\frac{n}{\sigma^2}\varepsilon} = \Delta + \frac{\varepsilon}{2} - \frac{\sigma^2 \ln\left(\frac{L_2\lambda_i'}{2L_1\lambda_i}\right)}{n\varepsilon}$$

and

$$s_i(n) = \frac{\frac{n}{2\sigma^2}(2\Delta+\varepsilon)\varepsilon - \ln\left(\frac{L_2\lambda_i'}{L_1\lambda_i}\right)}{\frac{n}{\sigma^2}\varepsilon} = \Delta + \frac{\varepsilon}{2} - \frac{\sigma^2 \ln\left(\frac{L_2\lambda_i'}{L_1\lambda_i}\right)}{n\varepsilon},$$

hence,

$$\lim_{n\to\infty} r_i(n) = \lim_{n\to\infty} s_i(n) = \Delta + \frac{\varepsilon}{2} \text{ for any } L_1, L_2, \lambda_i, \lambda_i', \sigma, \delta, \varepsilon.$$

So, $\lim_{n\to\infty} t_i(n) = \Delta + \frac{\varepsilon}{2}$, which completes the proof.

Now we give the proof of Theorem 1.5.1.

<u>Proof</u> (of Theorem 1.5.1.):

Now, $\bar{X}_{in} \sim N(\theta_i, \frac{\sigma^2}{n})$. For $|\theta_i - \theta_0| \le \Delta$, let

$$g(\theta_i) = E_{\theta_i}[\delta_i^*(\bar{X}_{in})] = Pr[-t_i(n) \le \bar{X}_{in} - \theta_0 \le t_i(n)]$$

$$= \Phi\left(\frac{t_i(n)+\theta_0-\theta_i}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{-t_i(n)+\theta_0-\theta_i}{\sigma/\sqrt{n}}\right).$$

If we recall that $g(\theta_i+\theta_0) = g(\theta_0-\theta_i)$, and g is decreasing for $\theta_i > \theta_0$, then $\inf\limits_{|\theta_i-\theta_0| \leq \Delta} g(\theta_i) = g(\theta_0+\Delta) = g(\theta_0-\Delta)$. Hence,

$$\inf_{|\theta_i-\theta_0| \leq \Delta} E_{\theta_i}[\delta_i^*(\bar{X}_{in})] = \Phi\left(\frac{t_i(n)-\Delta}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{-t_i(n)-\Delta}{\sigma/\sqrt{n}}\right).$$

So,

$$\lim_{n\to\infty} \inf_{|\theta_i-\theta_0| \leq \Delta} E_{\theta_i}[\delta_i^*(\bar{X}_{in})] = \Phi(\infty)-\Phi(-\infty) = 1.$$

Similarly,

$$\sup_{|\theta_i-\theta_0| \geq \Delta+\varepsilon} g(\theta_i) = g(\theta_0+\Delta+\varepsilon) = g(\theta_0-\Delta-\varepsilon).$$

Hence,

$$\lim_{n\to\infty} \sup_{|\theta_i-\theta_0| \geq \Delta+\varepsilon} E_{\theta_i}[\delta_i^*(\bar{X}_{in})] = \lim_{n\to\infty}\left[\Phi\left(\frac{t_i(n)-(\Delta+\varepsilon)}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{-t_i(n)-(\Delta+\varepsilon)}{\sigma/\sqrt{n}}\right)\right]$$

$$= \Phi(-\infty)-\Phi(-\infty) = 0.$$

This completes the proof of Theorem 1.5.1.

<u>Remark</u>: If θ_0 is unknown, Theorem 1.5.1. becomes

$$\lim_{n\to\infty} \inf_{|\theta_i-\theta_0| \leq \Delta} E_{\theta_i-\theta_0}[\delta_i^0(\bar{X}_{in}-\bar{X}_{0n})] = 1$$

and

$$\lim_{n\to\infty} \sup_{|\theta_i-\theta_0|\geq\Delta+\varepsilon} E_{\theta_i-\theta_0}[\delta_i^0(\bar{X}_{in}-\bar{X}_{0n})] = 0 .$$

The proof is similar to that of Theorem 1.5.1 and hence it is omitted.

<u>Theorem 1.5.2.</u> $\lim_{n\to\infty} \sup_{\tau\in\Gamma} r(\tau,\underset{\sim}{\delta}^*) = 0$, where $\underset{\sim}{\delta}^* = (\delta_1^*,\ldots,\delta_k^*)$ is the Γ minimax rule we found in Theorem 1.3.3. with x_i replaced by $\bar{x}_{in}$.

<u>Proof</u>: $\sup_{\tau\in\Gamma} r(\tau,\underset{\sim}{\delta}^*) \leq \sum_{i=1}^{k} \sup_{\tau\in\Gamma} r^{(i)}(\tau,\delta_i^*)$.

Now,

$$\sup_{\tau\in\Gamma} r^{(i)}(\tau,\delta_i^*) = \sup_{\tau\in\Gamma}\{\int_{|\theta_i-\theta_0|\leq\Delta} L_1(1-E_{\theta_i}[\delta_i^*(\bar{X}_{in})]d\tau(\theta_i)$$

$$+ \int_{|\theta_i-\theta_0|\geq\Delta+\varepsilon} L_2 E_{\theta_i}[\delta_i^*(\bar{X}_{in})]d\tau(\theta_i)$$

$$\leq L_1\lambda_i(1 - \inf_{|\theta_i-\theta_0|\leq\Delta} E_{\theta_i}[\delta_i^*(\bar{X}_{in})]) + L_2\lambda_i' \sup_{|\theta_i-\theta_0|\geq\Delta+\varepsilon} E_{\theta_i}[\delta_i^*(\bar{X}_{in})],$$

then

$$\lim_{n\to\infty} \sup_{\tau\in\Gamma} r(\tau,\underset{\sim}{\delta}^*) \leq \sum_{i=1}^{k} \lim_{n\to\infty} \sup_{\tau\in\Gamma} r^{(i)}(\tau,\delta_i^*)$$

$$\leq L_1\lambda_i(1-\lim_{n\to\infty} \inf_{|\theta_i-\theta_0|\leq\Delta} E_{\theta_i}[\delta_i^*(\bar{X}_{in})])$$

$$+ L_2\lambda_i' \lim_{n\to\infty} \sup_{|\theta_i-\theta_0|\geq\Delta+\varepsilon} E_{\theta_i}[\delta_i^*(\bar{X}_{in})] = 0 .$$

Again, when θ_0 is unknown, $\lim_{n\to\infty} \sup_{\tau\in\Gamma} r(\tau,\underset{\sim}{\delta}^0) = 0$ is also true.

<u>Theorem 1.5.3.</u> When θ_0 is known, the Γ-minimax rule $\underset{\sim}{\delta}^* = (\delta_1^*,\ldots,\delta_k^*)$ is admissible.

Proof: Let τ^* be as defined in the proof of Theorem 1.3.3., then, any Bayes rule $\hat{\underset{\sim}{\delta}} = (\hat{\delta}_1, \ldots, \hat{\delta}_k)$ of τ^* is of the form

$$\hat{\delta}_i(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } L_2\lambda_i[f_{\theta_0-\Delta-\varepsilon}(x_i)+f_{\theta_0+\Delta+\varepsilon}(x_i)] \\ & \qquad < L_1\lambda_i[f_{\theta_0-\Delta}(x_i)+f_{\theta_0+\Delta}(x_i)] \\ \nu_i(\underset{\sim}{x}) & \text{if } \qquad = \\ 0 & \text{if } \qquad > \end{cases}$$

However,

$$\{x_i | L_2\lambda_i[f_{\theta_0-\Delta-\varepsilon}(x_i)+f_{\theta_0+\Delta+\varepsilon}(x_i)] = L_1\lambda_i[f_{\theta_0-\Delta}(x_i)+f_{\theta_0+\Delta}(x_i)]\}$$

$$, \qquad \subseteq \{t_i, -t_i\} ,$$

and

$$P[X_i = \pm t_i] = 0 \text{ since } X_i \sim N(\theta_i, \sigma^2).$$

This shows that the Bayes rule of τ^* is unique up to equivalence. It follows that all Bayes rules of τ^* are admissible (Ferguson p. 60 [1967]). Particularly, $\underset{\sim}{\delta}^*$ (with $\nu_i(\underset{\sim}{x}) = 1$) is admissible.

When θ_0 is unknown, the Γ-minimax rule $\underset{\sim}{\delta}^0$ is also admissible. To prove this, we need to consider the generalized Bayes rule.

Definition 1.5.1. If τ is a measure on Θ, and if δ_0 satisfies

$$\inf_{\delta} \int_{\Theta} L(\theta,\delta(x))f(x|\theta)d\tau(\theta) = \int_{\Theta} L(\theta,\delta_0(x))f(x|\theta)d\tau(\theta) < \infty ,$$

then δ_0 is said to be a generalized Bayes rule for τ.

Remark: δ_0 is a generalized Bayes rule can't guarantee $r(\tau, \delta_0) < \infty$.

Lemma 1.5.2. If $|L(\theta,a)| \leq M$ for some constant M, then δ_0 is a generalized Bayes rule-> for all δ, $\int_{\Theta} [R(\theta,\delta)-R(\theta.\theta_0)]d\tau(\theta) \geq 0$.

Proof: For any δ,

$$\int_X \int_\Theta [L(\theta,\delta(x)) - L(\theta,\delta_0(x))] f(x|\theta) d\tau(\theta) dx \geq 0 .$$

But since $L(\theta,\delta(x)) - L(\theta,\delta_0(x))$ is bounded from below, by Fubini's Theorem,

$$\int_\Theta \int_X [L(\theta,\delta)-L(\theta,\delta_0)] f(x|\theta) dx d\tau(\theta) \geq 0 ,$$

i.e.,

$$\int_\Theta [R(\theta,\delta) - R(\theta,\delta_0)] d\tau(\theta) \geq 0 .$$

Definition 1.5.2. A generalized Bayes rule δ_0 wrt τ is unique up to equivalence iff for any rule δ,

$$\int_\Theta [R(\theta,\delta) - R(\theta,\delta_0)] d\tau(\theta) = 0$$

$$\Rightarrow R(\theta,\delta) = R(\theta,\delta_0) \ , \forall \ \theta.$$

Remark: Let δ_0 be a unique generalized Bayes rule according to definition 1.5.2., then if δ is any other generalized Bayes rule for τ, we have $R(\theta,\delta) = R(\theta,\delta_0)$ for all θ.

Lemma 1.5.3. If $L(\theta,a)$ is bounded and if the generalized Bayes rule δ_0 of τ is unique up to equivalence, then δ_0 is admissible.

Proof: Let δ be such that $R(\theta,\delta) \leq R(\theta,\delta_0)$ then

$$\int_\Theta [R(\theta,\delta) - R(\theta,\delta_0)] d\tau(\theta) \leq 0 .$$

By Lemma 1.5.2,

$$\int_\Theta [R(\theta,\delta) - R(\theta,\delta_0)] d\tau(\theta) = 0 .$$

Now by the uniqueness of δ_0, we get $R(\theta,\delta) = R(\theta,\delta_0)$, which completes the proof.

<u>Theorem 1.5.4</u>. When θ_0 is unknown, the Γ-minimax rule $\underset{\sim}{\delta}^0 = (\delta_1^0,\ldots,\delta_k^0)$ is admissible in D_1.

<u>Proof</u>: Let $\tau = (\tau_0, \omega_{\theta_0})$ be the measure on Θ such that τ_0 is Lebesque measure on Θ_0, and with θ_0 given, $\theta_1,\theta_2,\ldots,\theta_k$ are independent, such that

$$P_{\omega_{\theta_0}}[\theta_i = \theta_0+\Delta|\theta_0] = P_{\omega_{\theta_0}}[\theta_i = \theta_0-\Delta|\theta_0] = \frac{\lambda_i}{2}$$

$$P_{\omega_{\theta_0}}[\theta_i = \theta_0+\Delta+\varepsilon|\theta_0] = P_{\omega_{\theta_0}}[\theta_i = \theta_0-\Delta-\varepsilon|\theta_0] = \frac{\lambda_i'}{2}$$

$$P_{\omega_{\theta_0}}[\theta_i = \theta_0+\Delta+\frac{\varepsilon}{2}|\theta_0] = 1 - \lambda_i - \lambda_i' ,$$

then for any $\underset{\sim}{\delta} = (\delta_1,\ldots,\delta_k) \in D_1$,

$$\int_{\Theta} L(\underset{\sim}{\theta},\underset{\sim}{\delta}(\underset{\sim}{x}))f(\underset{\sim}{x}|\underset{\sim}{\theta})d\tau(\underset{\sim}{\theta})$$

$$= \sum_{i=1}^{k} \int_{-\infty}^{\infty} \frac{\lambda_i}{2} L_1(1-\delta_i(x_0,x_i))[f_{\theta_0+\Delta}(x_i)+f_{\theta_0-\Delta}(x_i)]f_{\theta_0}(x_0)$$

$$+ \frac{\lambda_i'}{2} L_2\ \delta_i(x_0,x_i)[f_{\theta_0+\Delta+\varepsilon}(x_i)+f_{\theta_0-\Delta-\varepsilon}(x_i)]f_{\theta_0}(x_0)\ d\theta_0$$

$$= \sum_{i=1}^{k} \int_{-\infty}^{\infty} \frac{\lambda_i}{2} L_1[f_{\theta_0+\Delta}(x_i)+f_{\theta_0-\Delta}(x_i)]f_{\theta_0}(x_0)d\theta_0$$

$$+ \sum_{i=1}^{k} \delta_i(x_0,x_i)\left\{\frac{\lambda_i'}{2\sqrt{2}} L_2[f_0(\frac{x_i-x_0+\Delta+\varepsilon}{\sqrt{2}}) + f_0(\frac{x_i-x_0-\Delta-\varepsilon}{\sqrt{2}})] - \right.$$

$$\left.\frac{\lambda_i}{2\sqrt{2}} L_1[f_0(\frac{x_i-x_0+\Delta}{\sqrt{2}}) + f_0(\frac{x_i-x_0-\Delta}{\sqrt{2}})]\right\}.$$

Hence, if $\delta_i^0(x_0,x_i)$ is defined by (1.4.6), then

$\delta^0 = (\delta_1^0,\ldots,\delta_k^0)$ is a generalized Bayes rule wrt τ. Now, if

$$\int_\Theta R(\underset{\sim}{\theta},\underset{\sim}{\delta}) - R(\underset{\sim}{\theta},\underset{\sim}{\delta}_0)d\tau(\underset{\sim}{\theta}) = 0 \ ,$$

i.e.,

$$\int_X \int_\Theta [L(\underset{\sim}{\theta},\underset{\sim}{\delta}(\underset{\sim}{x})) - L(\underset{\sim}{\theta},\underset{\sim}{\delta}^0(\underset{\sim}{x}))]f(\underset{\sim}{x}|\underset{\sim}{\theta})d\tau(\underset{\sim}{\theta})d\underset{\sim}{x} = 0$$

by Fubini's Theorem, so

$$\sum_{i=1}^{k}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\delta_i(x_0,x_i)-\delta_i^0(x_0,x_i)]\Big\{\frac{\lambda_i'}{2}L_2[f_0(\frac{x_i-x_0+\Delta+\varepsilon}{\sqrt{2}})+f_0(\frac{x_i-x_0-\Delta-\varepsilon}{\sqrt{2}})]$$

$$-\frac{\lambda_i}{2}L_2[f_0(\frac{x_i-x_0+\Delta}{\sqrt{2}})+f_0(\frac{x_i-x_0-\Delta}{\sqrt{2}})]\Big\}dx_i dx_0 = 0 \ .$$

Hence,

$$\delta_i(x_0,x_i) = \begin{cases} 1 & \text{if } L_2\,\lambda_i'[\phi(\frac{x_i-x_0+\Delta+\varepsilon}{\sqrt{2}\,\sigma}) + (\frac{x_i-x_0-\Delta-\varepsilon}{\sqrt{2}\,\sigma})] \\ & \quad < L_1\,\lambda_i[\phi(\frac{x_i-x_0+\Delta}{\sqrt{2}\,\sigma}) + (\frac{x_i-x_0-\Delta}{\sqrt{2}\,\sigma})] \\ \nu_i(x_0,x_i) & \text{if } = \\ 0 & \text{if } > \end{cases} \quad \text{a.e.}$$

Again,

$$\{(x_o,x_i)|L_2\lambda_i'[\phi(\frac{x_i-x_0+\Delta+\varepsilon}{\sqrt{2}\,\sigma}) + \phi(\frac{x_i-x_0-\Delta-\varepsilon}{\sqrt{2}\,\sigma})]$$

$$= L_1\lambda_i[\phi(\frac{x_i-x_o+\Delta}{\sqrt{2}\,\sigma}) + \phi(\frac{x_i-x_0-\Delta}{\sqrt{2}\,\sigma})] \subseteq \{x_i-x_0 = \pm\, t_i\}$$

and

$$P[X_i-X_0 = \pm\, t_i] = 0 \ , \quad \text{since} \quad \binom{X_i}{X_0} \sim N\left(\binom{\theta_i}{\theta_0} , \begin{pmatrix}\sigma^2 & 0\\ 0 & \sigma^2\end{pmatrix}\right) ,$$

So,

$$\delta_i(x_i, x_0) = \delta_i^0(x_i, x_0) \text{ a.e. }.$$

Then,

$$R(\underset{\sim}{\theta}, \underset{\sim}{\delta}) = R(\underset{\sim}{\theta}, \underset{\sim}{\delta}^0).$$

By Lemma 1.5.3., $\underset{\sim}{\delta}^0$ is admissible. This proves Theorem 1.5.4.

When θ_0 is unknown, we have restricted the decision rules to the class D_1. It is quite natural and reasonable for us to do this. However, we may still like to know:

a. Is $\underset{\sim}{\delta}^0 = (\delta_1^0, \ldots, \delta_k^0)$ a Γ-minimax decision rule in D rather than only in D_1?

b. Is $\underset{\sim}{\delta}^0$ admissible in D?

We leave these questions as open for further research.

1.6 Relaxing the assumption of normality

As was remarked in Section 1.3, the assumption of normality is somewhat restrictive to our problem. In this section, we will investigate some more general distributions for which Γ-minimax rules exist.

Theorem 1.6.1. Assume θ_0 is known. Let

$$A_i = \{x_i \mid \lambda_i' L_2[f_{\theta_0+\Delta+\varepsilon}(x_i) + f_{\theta_0-\Delta-\varepsilon}(x_i)] \le \lambda_i L_1[f_{\theta_0+\Delta}(x_i) + f_{\theta_0-\Delta}(x_i)]\}.$$

Let

$g(\theta_i) = E_{\theta_i}[I_{A_i}(X_i)]$ where $X_i \sim f_{\theta_i}(x)$. If $g(\theta_i+\theta_0) = g(\theta_0-\theta_i)$ and $g(\theta_i)$ is decreasing for $\theta_i > \theta_0$, then $\underset{\sim}{\delta} = (\delta_1, \ldots, \delta_k)$ is a Γ-minimax rule where $\delta_i(\underset{\sim}{x}) = I_{A_i}(x_i)$ for $i=1,\ldots,k$.

Proof: Let τ^* be defined as in the proof of Theorem 1.3.3., then the Bayes rule for the i^{th}-component problem wrt τ^* is given by $\delta_i(x_i) = I_{A_i}(x_i)$ for $i=1,2,\ldots,k$.

Now, since $g(\theta_i+\theta_0) = g(\theta_0-\theta_i)$ and $g(\theta_i)$ is decreasing for $\theta_i > \theta_0$, so

$$\sup_{|\theta_i-\theta_0|\geq\Delta+\varepsilon} g(\theta_i) = g(\theta_0+\Delta+\varepsilon) = g(\theta_0-\Delta-\varepsilon)$$

and

$$\inf_{|\theta_i-\theta_0|\leq\Delta} g(\theta_i) = g(\theta_0+\Delta) = g(\theta_0-\Delta) \ ,$$

i.e., δ_i satisfies (1.3.1). Hence by Theorem 1.3.1., $\underset{\sim}{\delta}$ is a Γ-minimax rule.

The following example applies Theorem 1.6.1. to select some binomial populations with large entropy.

Definition 1.6.1. For a binomial distribution $b(n,p)$, its entropy is defined as

$$\psi(p) = -\ [p\,\ell n p+(1-p)\ell n(1-p)] \ .$$

Note that $\psi(p)$ is associated with the uncertainty or randomness of that population. The larger the $\psi(p)$, the stronger the randomness.

Example 1.6.1.: Suppose $\Pi_1,\Pi_2,\ldots,\Pi_k$ are k independent binomial populations with $\Pi_i \sim b(n,p_i)$. We define

$$\Pi_i \text{ is positive iff } \psi(p_i) \geq \beta + \varepsilon$$

and $$\Pi_i \text{ is negative iff } \psi(p_i) \leq \beta \ .$$

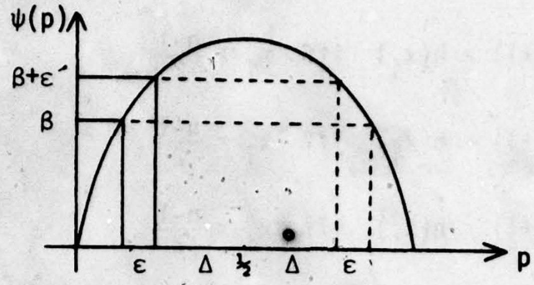


Figure 2. Graph of $\psi(p)$.

Equivalently, we can say that

Π_i is positive iff $|p_i - \frac{1}{2}| \le \Delta$

and Π_i is negative iff $|p_i - \frac{1}{2}| \ge \Delta + \varepsilon$,

where $\psi(\frac{1}{2} + \Delta) = \beta + \varepsilon'$ and $\psi(\frac{1}{2} + \Delta + \varepsilon) = \beta$.

It is seen that $\theta_0 = \frac{1}{2}$ in this problem. Let Γ be given by (1.2.2) and the loss given by (1.2.1) with $\theta_0 = \frac{1}{2}$ and $\theta_i = p_i$, then

$$A_i = \{x_i | L_2\lambda_i'[p_{\frac{1}{2}-\Delta-\varepsilon}(x_i)+p_{\frac{1}{2}+\Delta+\varepsilon}(x_i)] \le L_1\lambda_i[p_{\frac{1}{2}+\Delta}(x_i)+p_{\frac{1}{2}-\Delta}(x_i)]\}$$

where $p_\theta(x) = \binom{n}{\theta}\theta^x(1-\theta)^{n-x}$, we find $x_i \in A_i$ iff

$$\frac{L_2\lambda_i'}{L_1\lambda_i}\ \frac{(\frac{1}{2}-\Delta-\varepsilon)^{x_i}(\frac{1}{2}+\Delta+\varepsilon)^{n-x_i}+(\frac{1}{2}+\Delta+\varepsilon)^{x_i}+(\frac{1}{2}-\Delta-\varepsilon)^{n-x_i}}{(\frac{1}{2}-\Delta)^{x_i}(\frac{1}{2}+\Delta)^{n-x_i}+(\frac{1}{2}+\Delta)^{x_i}+(\frac{1}{2}-\Delta)^{n-x_i}} = h(x_i) \le 1 .$$

Also, it is obvious that $h(\frac{n}{2} - x_i) = h(\frac{n}{2} + x_i)$. After some messy computation, we get

$$h(x_i+1)-h(x_i) = c(\tfrac{1}{2}-\Delta-\varepsilon)^{x_i}(\tfrac{1}{2}+\Delta+\varepsilon)^{n-x_i-1}\{\varepsilon(\tfrac{1}{2}-\Delta)^{x_i}(\tfrac{1}{2}+\Delta)^{n-x_i-1}$$

$$[(\frac{\frac{1}{2}+\Delta+\varepsilon}{\frac{1}{2}-\Delta-\varepsilon}\cdot\frac{\frac{1}{2}+\Delta}{\frac{1}{2}-\Delta})^{2x_i+1-n}-1]+(2\Delta+\varepsilon)(\tfrac{1}{2}+\Delta)^{x_i}(\tfrac{1}{2}-\Delta)^{n-x_i-1}$$

$$[(\frac{\frac{1}{2}-\Delta}{\frac{1}{2}+\Delta}\cdot\frac{\frac{1}{2}+\Delta+\varepsilon}{\frac{1}{2}-\Delta-\varepsilon})^{2x_i+1-n}-1\ ]\},$$

then

$$h(x_i+1) > h(x_i) \quad \text{iff} \quad x_i > \frac{n-1}{2}$$

$$h(x_i+1) = h(x_i) \quad \text{iff} \quad x_i = \frac{n-1}{2}$$

$$h(x_i+1) < h(x_i) \quad \text{iff} \quad x_i < \frac{n-1}{2} .$$

Hence,

$$A_i = \{x_i | h(x_i) \le 1\} = \{x_i \mid \frac{n}{2} - m_i \le x_i \le \frac{n}{2} + m_i\} ,$$

for some integer m_i. It follows that

$$g(p_i) = E_{p_i}[I_{A_i}(X_i)] = P[\frac{n}{2} - m_i \le X_i \le \frac{n}{2} + m_i] ,$$

$$= \begin{cases} \displaystyle\sum_{x_i = \frac{n}{2} - m_i}^{\frac{n}{2} + m_i} \binom{n}{x_i} p_i^{x_i}(1-p_i)^{n-x_i} & \text{if } n \text{ is even} \\ \displaystyle\sum_{x_i = \frac{n+1}{2} - m_i}^{\frac{n-1}{2} + m_i} \binom{n}{x_i} p_i^{x_i}(1-p_i)^{n-x_i} & \text{if } n \text{ is odd} . \end{cases}$$

$$= g(1-p_i) ,$$

i.e., $g(\frac{1}{2}+p_i) = g(\frac{1}{2}- p_i)$. Since

$$\binom{n}{x_i} p_i^{x_i}(1-p_i)^{n-x_i} = (1-p_i)^n \binom{n}{x_i} e^{x_i \ln(\frac{p_i}{1-p_i})}$$

and $\ln(\frac{p_i}{1-p_i})$ is increasing in p_i, so by Corollary 1.3.1, $g(p_i)$ is decreasing for $p_i > \frac{1}{2}$ and increasing for $p_i < \frac{1}{2}$.

Now, by Theorem 1.6.1., $\underset{\sim}{\delta} = (\delta_1,\ldots,\delta_k)$ with $\delta_i(x_i) = I_{[\frac{n}{2} - m_i, \frac{n}{2} + m_i]}(x_i)$ is a Γ-minimax rule.

Definition 1.6.2. $X \sim f_\theta(x)$ is said to be a PF density (pólya frequency) if $f_\theta(x) = f(x-\theta)$ is TP.

Theorem 1.6.2. (Karlin) If $X \sim f_\theta(x)$ is a PF density and $f(x) = f(-x)$, then $|X|$ is TP (hence TP_2) for $\theta > 0$.

Remark: The density of $|X|$ is $[f_\theta(x)+f_\theta(-x)]\, I_{[0,\infty]}(x)$, so by Theorem 1.6.2, we can assert that

$$\frac{f(x-\theta_2) + f(x+\theta_2)}{f(x-\theta_1) + f(x+\theta_1)} \tag{1.6.1}$$

is increasing in x for $x > 0$ and $\theta_2 > \theta_1$.

Theorem 1.6.3. If X_i has a PF density $f_{\theta_i}(x) = f(x-\theta_i)$ and $f(x)=f(-x)$, then the assumptions of Theorem 1.6.1 are satisfied.

Proof: We need to show $g(\theta_i+\theta_0) = g(\theta_0-\theta_i)$ and g is decreasing for $\theta_i > \theta_0$. Let $y_i = x_i-\theta_0$, then

$$A_i = \{y_i+\theta_0 \mid \frac{L_2\lambda_i^*}{L_1\lambda_i} \frac{f(y_i+\Delta+\varepsilon) + f(y_i-\Delta-\varepsilon)}{f(y_i+\Delta) + f(y_i-\Delta)} \le 1\} .$$

Let

$$h(y_i) = \frac{f(y_i+\Delta+\varepsilon) + f(y_i-\Delta-\varepsilon)}{f(y_i+\Delta) + f(y_i-\Delta)} ,$$

then

$$f(x) = f(-x) \Rightarrow h(y_i) = h(-y_i) .$$

Also, from (1.6.1), h is increasing in y_i for $y_i > 0$,

so $A_i = \{y_i+\theta_0 \mid -t_i \le y_i \le t_i\}$. Then

$$g(\theta_i) = E_{\theta_i}[I_{A_i}(X_i)] = P[-t_i+\theta_0 \le Z+\theta_i \le t_i+\theta_0] ,$$

where $Z \sim f_{\theta_i=0}(x)$. Since $Z \sim -Z$, hence

$$g(\theta_i+\theta_0) = g(\theta_0-\theta_i) \ .$$

Now, by the remark of Corollary 1.3.1., we get g is decreasing in θ_i for $\theta_i > \theta_0$. This completes the proof.

Example 1.6.2.: If $f_{\theta_i}(x) = \frac{c_i}{2} e^{-c_i|x_i-\theta_i|}$ for $i=1,2,\ldots,k$, where c_i's are known constants, then the Γ-minimax rule is $\underset{\sim}{\delta} = (\delta_1,\ldots,\delta_k)$ with

$$\delta_i(x_i) = \begin{cases} 1 & \text{if } \dfrac{\lambda_i L_2 \, [e^{-c_i|x_i-\theta_0-\Delta-\varepsilon|} + e^{-c_i|x_i-\theta_0+\Delta+\varepsilon|}]}{\lambda_i L_1 \, [e^{-c_i|x_i-\theta_0-\Delta|} + e^{-c_i|x_i-\theta_0+\Delta|}]} \leq 1 \\ 0 & \qquad > \end{cases}$$

Proof: The result follows directly from Theorem 1.6.3.

1.7 Bayes rule and the minimax rule for selecting populations close to a control

In Section 1.3, we assumed that partial prior informations about Θ are known and that they are summarized in the class Γ. In this section, we will consider two extreme cases, namely, either complete information or no information about Θ is known. Correspondingly, we are looking for the Bayes rule or minimax rule. The problem will be treated under the assumption that θ_0 is unknown.

The following lemma may have been used by many people implicitly, but it is worth stating it out explicitly.

Lemma 1.7.1. If (Θ,D,L) is a decision problem and if for any $\underset{\sim}{\delta} \in D$, $\underset{\sim}{\delta} = (\delta_1,\ldots,\delta_k)$ and $L(\theta,\underset{\sim}{\delta}) = \sum_{i=1}^{k} L^{(i)}(\theta,\delta_i)$ [i.e., the loss is additive],

then for any prior distribution τ on Θ, $\underset{\sim}{\delta}^0$ is a Bayes rule of τ if δ_i^0 is a Bayes rule of τ for the i^{th} component problem.

<u>Proof</u>: $r(\tau,\underset{\sim}{\delta}) = \sum_{i=1}^{k} r^{(i)}(\tau,\delta_i) \geq \sum_{i=1}^{k} r^{(i)}(\tau,\delta_i^0) = r(\tau,\underset{\sim}{\delta}^0)$.

Let us assume $\theta_i \sim \tau_i(\theta) \sim N(\alpha_i,\beta_i^2)$ and θ_i's are independent ($0\leq i\leq k$), then since $X_i|\theta_i \sim N(\theta_i,\sigma^2)$, we get

$$\theta_i|X_i \sim N\left(\frac{\alpha_i\sigma^2 + x_i\beta_i^2}{\sigma^2 + \beta_i^2}, \frac{\sigma^2\,\beta_i^2}{\sigma^2 + \beta_i^2}\right) = N(a_i, b_i^2) .$$

The Bayes rule of the i^{th} component problem is to minimize

$$\int_X [L_1(1-\delta_i(\underset{\sim}{x}))\int_{|\theta_i-\theta_0|\leq\Delta} d\tau(\underset{\sim}{\theta}|\underset{\sim}{x}) + L_2\delta_i(\underset{\sim}{x})\int_{|\theta_i-\theta_0|\geq\Delta+\varepsilon} d\tau(\underset{\sim}{\theta}|\underset{\sim}{x})]m(\underset{\sim}{x})d\underset{\sim}{x} , \tag{1.7.1}$$

where

$$m(\underset{\sim}{x}) = \int_\Theta \prod_{i=0}^{k} f_{\theta_i}(x_i)d\tau_i(\theta_i) = \prod_{i=0}^{k} \int_{-\infty}^{\infty} f_{\theta_i}(x_i)d\tau_i(\theta_i) = \prod_{i=1}^{k} m_i(x_i)$$

and

$$d\tau(\underset{\sim}{\theta}|\underset{\sim}{x}) = \frac{\prod_{i=0}^{k} f_{\theta_i}(x_i)d\tau_i(\theta_i)}{m(\underset{\sim}{x})} = \prod_{i=0}^{k} \frac{f_{\theta_i}(x_i)d\tau_i(\theta_i)}{m_i(x_i)} = \prod_{i=1}^{k} d\tau_i(\theta_i|x_i) .$$

Hence, (1.7.1) reduces to

$$L_1\int_X\int_{|\theta_i-\theta_0|\leq\Delta} d\tau(\underset{\sim}{\theta}|\underset{\sim}{x})m(\underset{\sim}{x})d\underset{\sim}{x} + \int_X [L_2\int_{|\theta_i-\theta_0|\geq\Delta+\varepsilon} d\tau(\theta_0|x_0)d\tau(\theta_i|x_i)$$

$$- L_1\int_{|\theta_i-\theta_0|\leq\Delta} d\tau(\theta_0|x_0)d\tau(\theta_i|x_i)]\delta_i(\underset{\sim}{x}) \prod_{\substack{j\neq 0\\ j\neq i}}^{k} d\tau(\theta_j|x_j)m(\underset{\sim}{x})d\underset{\sim}{x} .$$

So the Bayes rule is

$$\delta_i^B(x_0,x_i) = \begin{cases} 1 & \text{if } L_2P[|\theta_i-\theta_0|\geq\Delta+\varepsilon|x_0,x_i] \leq L_1P[|\theta_i-\theta_0|\leq\Delta|x_0,x_i] \\ 0 & \qquad\qquad > \end{cases} .$$

But $\theta_i-\theta_0|x_0,x_i \sim N(a_i-a_0,\ b_i^2+b_0^2)$, thus we have

$$\delta_i^B(x_0,x_i)=\begin{cases}1 & \text{if } L_2\left[1-\Phi\left(\dfrac{\Delta+\varepsilon+a_0-a_i}{\sqrt{b_0^2+b_i^2}}\right)+\Phi\left(\dfrac{-\Delta-\varepsilon+a_0-a_i}{\sqrt{b_0^2+b_i^2}}\right)\right] \\ & \le L_1\left[\Phi\left(\dfrac{\Delta+a_0-a_i}{\sqrt{b_0^2+b_i^2}}\right)-\Phi\left(\dfrac{-\Delta+a_0-a_i}{\sqrt{b_0^2+b_i^2}}\right)\right] \\ 0 & >\end{cases}$$

$$=\begin{cases}1 & \text{if } h(y_i)=\dfrac{L_2[1-\Phi(y_i+\Delta'+\varepsilon')+\Phi(y_i-\Delta'-\varepsilon')]}{\Phi(y_i+\Delta')-\Phi(y_i-\Delta')}\le 1 \\ 0 & >\end{cases}$$

$$=\begin{cases}1 & \text{if } |y_i|\le s_i^B \\ 0 & >\end{cases}=\begin{cases}1 & \text{if } |a_i-a_0|\le t_i^B=s_i^B\sqrt{b_0^2+b_i^2} \\ 0 & >\end{cases} \tag{1.7.2}$$

where

$$y_i=\frac{a_0-a_i}{\sqrt{b_0^2+b_i^2}},\quad \Delta'=\frac{\Delta}{\sqrt{b_0^2+b_i^2}},\ \text{and}\ \varepsilon'=\frac{\varepsilon}{\sqrt{b_0^2+b_i^2}}.$$

The last equality holds because $h(y_i)=h(-y_i)$ and the fact that $h(y_i)$ is increasing for $y_i>0$. We have

<u>Theorem 1.7.1</u>. Assume θ_i has independent prior distribution $N(\alpha_i,\beta_i^2)$, $i=0,1,\ldots,k$, then the Bayes rule is $\underset{\sim}{\delta}^B=(\delta_1^B,\ldots,\delta_k^B)$ with δ_i^B defined by (1.7.2).

<u>Proof</u>: δ_i^B is the Bayes rule for i^{th} component problem, since $L = \sum_{i=1}^{k} L^{(i)}$. Hence Lemma 1.7.1 asserts that $\underset{\sim}{\delta}^B$ is a Bayes rule.

The following lemma is essential when we search for minimax rules.

<u>Lemma 1.7.2</u>. If $f_\theta(x) = f(x-\theta)$ is PF_2 and $f(x) = f(-x)$, then $f(y_0) \le f(x_0)$ if $y_0 \ge x_0 \ge 0$. Let $F(t) = \int_{-\infty}^{t} f(x)dx$, then for $t \ge 0$ and $\xi_2 \ge \xi_1 \ge 0$, we have

(i) $F(-t-\xi_1) + F(-t+\xi_1) \le F(-t-\xi_2) + F(-t+\xi_2)$

(ii) $F(t+\xi_1) - F(-t+\xi_1) \ge F(t+\xi_2) - F(-t+\xi_2)$.

<u>Proof</u>: (i) Let $\theta_0 = y_0 - x_0 \ge 0$, then $\dfrac{f_{\theta_0}(x)}{f_0(x)}$ is increasing in x. But when $x = \dfrac{\theta_0}{2}$,

$$\frac{f_{\theta_0}(\frac{\theta_0}{2})}{f_0(\frac{\theta_0}{2})} = \frac{f(-\frac{\theta_0}{2})}{f(\frac{\theta_0}{2})} = 1 ,$$

so $\dfrac{f_{\theta_0}(x)}{f_0(x)} \ge 1$ for all $x > \dfrac{\theta_0}{2} = \dfrac{y_0 - x_0}{2}$, hence

$$\frac{f_{\theta_0}(y_0)}{f_0(y_0)} \ge 1 \text{ , i.e., } f(y_0-\theta_0) = f(x_0) \ge f(y_0) .$$

Now,

$$F(-t-\xi_1) - F(-t-\xi_2) = \int_{-t-\xi_2}^{-t-\xi_1} f(x)dx = \int_0^{\xi_2-\xi_1} f(x-t-\xi_2)dx$$

$$\le \int_0^{\xi_2-\xi_1} f(x-t-\xi_1)dx = \int_{-t-\xi_1}^{-t-\xi_2} f(x)dx = F(-t-\xi_2) - F(-t-\xi_1)$$

(ii) Similar to (i).

Theorem 1.7.2. Let $\lambda_i = a$ and $\lambda_i' = 1-a$. Also, let $\underset{\sim}{\delta} = (\delta_1, \ldots, \delta_k)$, where $\delta_i(x_i - x_0) = I_{[t_i, t_i]}(x_i - x_0)$, be the corresponding Γ-minimax rule in D_1. If a is chosen so that

$$L_1[\Phi(\frac{-t_i+\Delta}{\sqrt{2}\,\sigma}) + \Phi(\frac{-t_i-\Delta}{\sqrt{2}\,\sigma})] = L_2[\Phi(\frac{t_i+\Delta+\varepsilon}{\sqrt{2}\,\sigma}) - \Phi(\frac{-t_i+\Delta+\varepsilon}{\sqrt{2}\,\sigma})] \,, \tag{1.7.3}$$

then $\underset{\sim}{\delta}$ is a minimax rule.

Proof: For $\underset{\sim}{\theta} \in \Theta_G(i)$,

$$\begin{aligned} R^{(i)}(\theta, \delta_i) &= L_1 \, P[|x_i - x_0| \geq t_i | \theta_0, \theta_i] \\ &= L_1 \, [\Phi(\frac{-t_i-(\theta_i-\theta_0)}{\sqrt{2}\,\sigma}) + \Phi(\frac{-t_i+(\theta_i-\theta_0)}{\sqrt{2}\,\sigma})] \\ &\leq L_1 \, [\Phi(\frac{-t_i-\Delta}{\sqrt{2}\,a}) + \Phi(\frac{-t_i+\Delta}{\sqrt{2}\,a})] \,, \end{aligned}$$

by Lemma 1.7.2(i). For $\underset{\sim}{\theta} \in \Theta_B(i)$,

$$\begin{aligned} R^{(i)}(\underset{\sim}{\theta}, \delta_i) &= L_2 \, [\Phi(\frac{t_i-(\theta_i-\theta_0)}{\sqrt{2}\,\sigma}) - \Phi(\frac{-t_i-(\theta_i-\theta_0)}{\sqrt{2}\,\sigma})] \\ &\leq L_2 \, [\Phi(\frac{t_i+\Delta+\varepsilon}{\sqrt{2}\,\sigma}) - \Phi(\frac{-t_i+\Delta+\varepsilon}{\sqrt{2}\,\sigma})] \end{aligned}$$

by Lemma 1.7.2(ii). And for $\underset{\sim}{\theta} \notin \Theta_B(i) \cup \Theta_G(i)$,

$$R^{(i)}(\underset{\sim}{\theta}, \delta_i) = 0 \,.$$

But in the proof of Theorem 1.4.1, we have shown

$$\liminf_{n\to\infty} r^{(i)}(\tau_n,\delta^0_{in}) \geq L_1 a(1-E_\Delta[\delta_i(Y_i)] + L_2(1-a)E_{\Delta+\varepsilon}[\delta_i(Y_i)]$$

$$= L_1 a[\Phi(\frac{-t_i+\Delta}{\sqrt{2}\,a}) + \Phi(\frac{-t_i-\Delta}{\sqrt{2}\,a})] + L_2(1-a)[\Phi(\frac{t_i+\Delta+\varepsilon}{\sqrt{2}\,a}) - \Phi(\frac{-t_i+\Delta+\varepsilon}{\sqrt{2}\,a})]$$

$$= L_1[\ (\frac{t_i+\Delta+\varepsilon}{\sqrt{2}\,a}) + \Phi(\frac{-t_i-\Delta}{\sqrt{2}\,a})]$$

$$= L_2[\ (\frac{t_i+\Delta+\varepsilon}{\sqrt{2}\,a}) - \Phi(\frac{-t_i+\Delta+\varepsilon}{\sqrt{2}\,a})]$$

$$\geq \sup_{\underset{\sim}{\theta}} R^{(i)}(\underset{\sim}{\theta},\delta_i)\ .$$

Hence,

$$\liminf_{n\to\infty} r(\tau_n,\underset{\sim}{\delta}^0_n) \geq \sum_{i=1}^{k} \liminf_{n\to\infty} r^{(i)}(\tau_n,\delta^0_{in})$$

$$\geq \sum_{i=1}^{k} \sup_{\underset{\sim}{\theta}} R^{(i)}(\underset{\sim}{\theta},\delta_i) \geq \sup_{\underset{\sim}{\theta}} \sum_{i=1}^{k} R^{(i)}(\underset{\sim}{\theta},\delta_i)$$

$$= \sup_{\underset{\sim}{\theta}} R(\underset{\sim}{\theta},\underset{\sim}{\delta})\ .$$

This proves $\underset{\sim}{\delta}$ is a minimax rule.

One may wonder the existence of such an a $(0<a<1)$, so that (1.7.3) holds. We will show that they do actually exist. We know t_i's are the positive roots of the equation

$$h_a(x) = \frac{L_2(1-a)}{L_1 a} e^{-\frac{1}{4\sigma^2}(2\Delta+\varepsilon)\varepsilon} \frac{\cosh(\frac{1}{2\sigma^2}(\Delta+\varepsilon)x)}{\cosh(\frac{1}{2\sigma^2}\Delta x)} - 1 = 0.$$

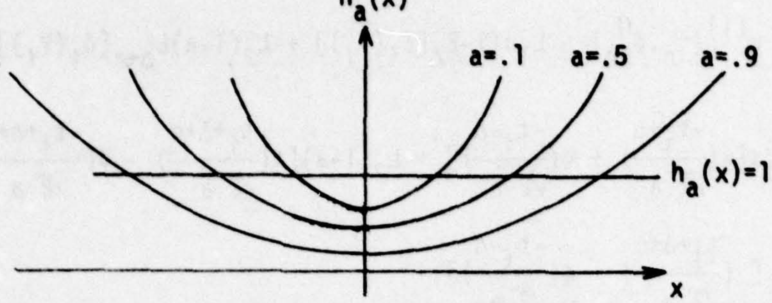


Figure 3. Graph of $h_a(x)$ for $a = .1$, $.5$ and $.9$.

One can see $h_a(x)$ is decreasing in a for fixed x, this implies t_i is increasing in a. And when $a \to 1$, we have $t_i \to \infty$, so

$$L_1[\Phi(\frac{-t_i+\Delta}{\sqrt{2}\,\sigma}) + \Phi(\frac{-t_i-\Delta}{\sqrt{2}\,\sigma})] \to 0$$

and

$$L_2[\Phi(\frac{t_i+\Delta+\varepsilon}{\sqrt{2}\,\sigma}) - \Phi(\frac{-t_i+\Delta+\varepsilon}{\sqrt{2}\,\sigma})] \to L_2 .$$

On the other hand, when $a \to 0$, $h_a(x)-1$ is positive for all x, so there exists some a_0 such that $t_i = 0$, then

$$L_1[\Phi(\frac{-t_i+\Delta}{\sqrt{2}\,\sigma}) + \Phi(\frac{-t_i-\Delta}{\sqrt{2}\,\sigma})] = L_1[\Phi(\frac{\Delta}{\sqrt{2}\,\sigma}) + \Phi(\frac{-\Delta}{\sqrt{2}\,\sigma})] = L_1$$

and

$$L_2[\Phi(\frac{t_i+\Delta+\varepsilon}{\sqrt{2}\,\sigma}) - \Phi(\frac{-t_i+\Delta+\varepsilon}{\sqrt{2}\,\sigma})] = 0 .$$

It is clear that

$$L_1[\Phi(\frac{-t_i+\Delta}{\sqrt{2}\,\sigma}) + \Phi(\frac{-t_i-\Delta}{\sqrt{2}\,\sigma})] - L_2[\Phi(\frac{t_i+\Delta+\varepsilon}{\sqrt{2}\,\sigma}) - \Phi(\frac{-t_i+\Delta+\varepsilon}{\sqrt{2}\,\sigma})]$$

is continuous in a, hence there exists a^* ($a_0 < a^* < 1$) such that (1.7.3) is true.

In the next section, a* will be found for some selected values of Δ and ε. $(L_1 = L_2 = 1)$.

1.8 Comparison among Bayes, Γ-minimax and minimax rules

When we face a decision problem, the prior information has a very important influence on our choices of the optimal rules. In general, one would use the Bayes rule if the prior distribution is known, use the Γ-minimax rule for incomplete prior information, and use the minimax rule when no prior information is available. The comparison of these rules will give us some idea about how far our decision is from the real optimal rule if the prior information we have is incorrect. In other words, we are interested in the robustness of each rule.

In this section, we make a thorough comparison among these rules in terms of Bayes risk, the maximum risk over Γ, and the overall maximum risk. Because the loss is assumed to be additive, the comparison is made for the 1^{st} component problem only. Sub-index i will be omitted from the notation and $\delta_B(\underset{\sim}{x}) = I_{[-t_B,t_B]}(a_1-a_0)$, $\delta_G(\underset{\sim}{x}) = I_{[-t_G,t_G]}(x_1-x_0)$, and $\delta_M(\underset{\sim}{x}) = I_{[-t_M,t_M]}(x_1-x_0)$ will mean Bayes rule, Γ-minimax rule, and the minimax rule, respectively. It is also understood that $d\tau_B(\theta_0,\theta_1) = d\tau_0(\theta_0)d\tau_1(\theta_1)$, where $\tau_i(\theta_i) \sim N(\alpha_i,\beta_i^2)$ is the prior and

$$a_i = \frac{\alpha_i\sigma^2 + x_i\beta_i^2}{\sigma^2 + \beta_i^2}, \quad \text{for } i=0,1. \quad \text{Also, } \lambda_1 = P_{\tau_B}[|\theta_1-\theta_0| \le \Delta],$$

$$\lambda_1' = P_{\tau_B}[|\theta_1-\theta_0| \ge \Delta+\varepsilon], \quad \underset{\sim}{x} = (x_0,x_1), \quad \text{and } \underset{\sim}{\theta} = (\theta_0,\theta_1)$$

will be used in this section. Now, the Bayes risk of the Bayes rule is

$r(\tau_B,\delta_B)$

$$= L_1 P[|a_1-a_0| > t_B,\ |\theta_1-\theta_0| \le \Delta]$$

$$+ L_2 P[|a_1-a_0| \le t_B,\ |\theta_1-\theta_0| \ge \Delta+\varepsilon] .$$

Since
$$\begin{pmatrix} x_i \\ \theta_i \end{pmatrix} \sim N\left(\begin{pmatrix} \alpha_i \\ \alpha_i \end{pmatrix}, \begin{pmatrix} \sigma^2 + \beta_i^2 & \beta_i^2 \\ \beta_i^2 & \beta_i^2 \end{pmatrix}\right)$$
for $i = 0,1$, then
$$\begin{pmatrix} a_1-a_0 \\ \theta_1-\theta_0 \end{pmatrix} \sim N\left(\begin{pmatrix} \alpha_1-\alpha_0 \\ \alpha_1-\alpha_0 \end{pmatrix}, \begin{pmatrix} \omega_0^2+\omega_1^2 & \omega_0^2+\omega_1^2 \\ \omega_0^2+\omega_1^2 & \omega_0^2+\omega_1^2 \end{pmatrix}\right),$$

where $\omega_i^2 = \dfrac{\beta_i^4}{\sigma^2+\beta_i^2}$, for $i = 0,1$. Let $d = \alpha_1-\alpha_0$,

$u^2 = \beta_0^2 + \beta_1^2$, $v^2 = \omega_0^2 + \omega_1^2$, and $\rho = \dfrac{v}{u}$, then

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} \dfrac{a_1-a_0-d}{v} \\ \dfrac{\theta_1-\theta_0-d}{u} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

Hence,

$$r(\tau_B,\delta_B) = L_1 P[(Z_1 > \frac{t_B-d}{v} \text{ or } Z_1 < \frac{-t_B-d}{v}) \text{ and } (\frac{-\Delta-d}{u} \le Z_2 \le \frac{\Delta-d}{u})]$$

$$+ L_2 P[(\frac{-t_B-d}{v} \le Z_1 \le \frac{t_B-d}{v}) \text{ and } (\frac{-\Delta-\varepsilon-d}{u} \ge Z_2 \text{ or } Z_2 \ge \frac{\Delta+\varepsilon-d}{u})]$$

$$= L_1[F(\frac{-t_B+d}{v}, \frac{\Delta-d}{u}; -\rho) - F(\frac{-t_B+d}{v}, \frac{-\Delta-d}{u}; -\rho)$$

$$+F(\frac{-t_B-d}{v}, \frac{\Delta-d}{u}; \rho) - F(\frac{-t_B-d}{v}, \frac{-\Delta-d}{u}; \rho)]$$

$$+ L_2[F(\frac{t_B-d}{v}, \frac{-\Delta-\varepsilon-d}{u}; \rho) - F(\frac{-t_B-d}{v}, \frac{-\Delta-\varepsilon-d}{u}; \rho)$$

$$+ F(\frac{t_B-d}{v}, \frac{-\Delta-\varepsilon+d}{u}; -\rho) - F(\frac{-t_B-d}{v}, \frac{-\Delta-\varepsilon+d}{u}; -\rho)]$$

where

$$F(x_0,y_0; \rho) = P[Z_1 \le x_0, Z_2 \le y_0] .$$

Similarly, we can compute

$$r(\tau,\delta_G) = L_2P[|X_1-X_0| \le t_G, |\Theta_1-\Theta_0| \ge \Delta+\varepsilon]$$

$$+ L_1P[|X_1-X_0| > t_G, |\Theta_1-\Theta_0| \le \Delta] .$$

Now,

$$\begin{pmatrix} X_1-X_0 \\ \Theta_1-\Theta_0 \end{pmatrix} \sim N\left(\begin{pmatrix} \alpha_1-\alpha_0 \\ \alpha_1-\alpha_0 \end{pmatrix}, \begin{pmatrix} \sigma^2+\beta_0^2+\sigma^2+\beta_1^2 & \beta_0^2+\beta_1^2 \\ \beta_0^2+\beta_1^2 & \beta_0^2+\beta_1^2 \end{pmatrix}\right)$$

$$= N\left(\begin{pmatrix} d \\ d \end{pmatrix}, \begin{pmatrix} r^2 & u^2 \\ u^2 & u^2 \end{pmatrix}\right),$$

where $r^2 = 2\sigma^2 + u^2$. Then

$$r(\tau_B,\delta_G) = L_1[F(\frac{-t_G+d}{r}, \frac{\Delta-d}{u}; -\rho') - F(\frac{-t_G+d}{r}, \frac{-\Delta-d}{u}; -\rho')$$

$$+ F(\frac{-t_G-d}{r}, \frac{\Delta-d}{u}; \rho') - F(\frac{-t_G-d}{r}, \frac{-\Delta-d}{u}; \rho')]$$

$$+ L_2[F(\frac{t_G-d}{r}, \frac{-\Delta-\varepsilon+d}{u}, -\rho') - F(\frac{-t_G-d}{r}, \frac{-\Delta-\varepsilon+d}{u}; -\rho')$$

$$+ F(\frac{t_G-d}{r}, \frac{-\Delta-\varepsilon-d}{u}; \rho') - F(\frac{-t_G-d}{r}, \frac{-\Delta-\varepsilon-d}{u}; \rho')] ,$$

where $\rho' = \frac{u}{r}$. Since δ_G and δ_M have the same form except for the constant t_G and t_M, so when we replace t_G by t_M in the above formula, we get $r(\tau_B,\delta_M)$.

The next thing we are going to compute is

$\sup_{\tau\in\Gamma} r^{(i)}(\tau,\delta)$, for $\delta = \delta_B$, δ_G and δ_M.

Now,

$$r^{(i)}(\tau,\delta) = \int_{|\theta_1-\theta_0|\le\Delta} L_1[1-E_{\underset{\sim}{\theta}}(\delta(\underset{\sim}{X}))]d\tau(\underset{\sim}{\theta})$$

$$+ \int_{|\theta_1-\theta_0|\ge\Delta+\varepsilon} L_2 E_{\underset{\sim}{\theta}}[\delta(\underset{\sim}{X})]d\tau(\underset{\sim}{\theta}) \ . \qquad (1.8.1)$$

Lemma 1.8.1.

$$\sup_{\tau\in\Gamma} r^{(1)}(\tau,\delta) = L_1(1- \inf_{|\theta_1-\theta_0|\le\Delta} E_{\underset{\sim}{\theta}}[\delta(\underset{\sim}{X})])\lambda_1 + L_2 \sup_{|\theta_1-\theta_0|\ge\Delta+\varepsilon} E_{\underset{\sim}{\theta}}[\delta(\underset{\sim}{X})]\lambda_1'$$

Proof: $\le$ is trivial. To prove the other inequality, let us consider two sequence $\{\underset{\sim}{\theta}_n\}_{n=1}^{\infty}$ and $\{\underset{\sim}{\theta}_n'\}_{n=1}^{\infty}$ such that $\underset{\sim}{\theta}_n \in \Theta_G(1)$ and $\underset{\sim}{\theta}_n' \in \Theta_B(1)$, and $E_{\underset{\sim}{\theta}_n}[\delta(\underset{\sim}{X})] \to \inf_{|\theta_1-\theta_0|\le\Delta} E_{\underset{\sim}{\theta}}[\delta(\underset{\sim}{X})]$ and $E_{\underset{\sim}{\theta}_n'}[\delta(\underset{\sim}{X})] \to \sup_{|\theta_1-\theta_0|\ge\Delta+\varepsilon} E_{\underset{\sim}{\theta}}[\delta(\underset{\sim}{X})]$.

If we define $\tau_n\in\Gamma$ as $P_{\tau_n}[\underset{\sim}{\theta}=\theta_n] = \lambda_1$, $P_{\tau_n}[\underset{\sim}{\theta}= \theta_n'] = \lambda_1'$,

and $P_{\tau_n}[\underset{\sim}{\theta} \notin \Theta_G(1) \cup \Theta_B(1)] = 1-\lambda_1-\lambda_1'$, then we have

$$\lim_{n\to\infty} r^{(1)}(\tau_n,\delta) = L_1(1- \inf_{|\theta_1-\theta_0|\le\Delta} E_{\underset{\sim}{\theta}}[\delta(\underset{\sim}{X})]\lambda_1 + L_2 \sup_{|\theta_1-\theta_0|\ge\Delta+\varepsilon} E_{\underset{\sim}{\theta}}[\delta(\underset{\sim}{X})]\lambda_1'.$$

Now, $\sup_{\tau\in\Gamma} r^{(1)}(\tau,\delta) \ge r^{(1)}(\tau_n,\delta)$ for all n, and if we take $\lim_{n\to\infty}$, we get the result.

From the above lemma and the proof of Theorem 1.4.1, we get

$$\sup_{\tau\in\Gamma} r^{(1)}(\tau,\delta_G) = L_1\lambda_1(1-E_\Delta[\delta_G(Y)]) + L_2\lambda_1' E_{\Delta+\varepsilon}[\delta_G(Y)] \ ,$$

where $Y = X_1 - X_0 \sim N(\theta_1 - \theta_0, 2\sigma^2)$. So,

$$\sup_{\tau\in\Gamma} r^{(1)}(\tau,\delta_G) = L_1\lambda_1 P[|X_1-X_0|>t_G|\theta_1-\theta_0=\Delta]$$

$$+ L_2\lambda_1' P[|X_1-X_0|\le t_G|\theta_1-\theta_0=\Delta+\varepsilon]$$

$$= L_1\lambda_1[\Phi(\frac{-t_G-\Delta}{\sqrt{2}\,\sigma}) + \Phi(\frac{-t_G+\Delta}{\sqrt{2}\,\sigma})] + L_2\lambda_1'[\Phi(\frac{t_G-\Delta-\varepsilon}{\sqrt{2}\,\sigma}) - \Phi(\frac{-t_G-\Delta-\varepsilon}{\sqrt{2}\,\sigma})] .$$

Hence we also get

$$\sup_{\tau\in\Gamma} r^{(1)}(\tau,\delta_M) = L_1\lambda_1[\Phi(\frac{-t_M-\Delta}{\sqrt{2}\,\sigma}) + \Phi(\frac{-t_M+\Delta}{\sqrt{2}\,\sigma})]$$

$$+ L_2\lambda_1'[\Phi(\frac{t_M-\Delta-\varepsilon}{\sqrt{2}\,\sigma}) - \Phi(\frac{-t_M-\Delta-\varepsilon}{\sqrt{2}\,\sigma})]$$

$$= (\lambda_1+\lambda_2)\, L_1[\Phi(\frac{-t_M-\Delta}{\sqrt{2}\,\sigma}) + \Phi(\frac{-t_M+\Delta}{\sqrt{2}\,\sigma})] .$$

To find $\sup_{\tau\in\Gamma} r^{(1)}(\tau,\delta_B)$ will need some more work. First note that

$a_1 - a_0|\theta_1,\theta_0 \sim N(\mu,\zeta^2)$, where

$$\mu = \frac{\beta_1^2\ \theta_1}{\beta_1^2+\sigma^2} - \frac{\beta_0^2\ \theta_0}{\beta_0^2+\sigma^2} + (\frac{\alpha_1\ \sigma^2}{\beta_1^2+\sigma^2} - \frac{\alpha_0\ \sigma^2}{\beta_0^2+\sigma^2})$$

and

$$\zeta^2 = \frac{\beta_1^4\ \sigma^2}{(\beta_1^2+\sigma^2)^2} + \frac{\beta_0^4\ \sigma^2}{(\beta_0+\sigma^2)^2} .$$

Then let

$$g(\mu) = E_\theta[\delta_B(\underset{\sim}{X})] = E_\mu[-t_B \le a_1-a_0 \le t_B]$$

$$= \Phi(\frac{t_B-\mu}{\zeta}) - \Phi(\frac{-t_B-\mu}{\zeta})$$

$$= \Phi(\frac{t_B+\mu}{\zeta}) - \Phi(\frac{-t_B+\mu}{\zeta}) ,$$

we find $g(\mu) = g(-\mu)$ and $g(\mu)$ is decreasing in $|\mu|$, by Lemma 1.7.2(ii). Now let us consider the following two cases.

(a) If $\beta_1^2 \neq \beta_0^2$, then $\frac{\beta_1^2}{\beta_1^2 + \sigma^2} \neq \frac{\beta_0^2}{\beta_0^2 + \sigma^2}$. And if we let

$\theta_1 = \theta_0 \to \pm \infty$, we have $|\mu| \to \infty$. So we get

$$\inf_{|\theta_1-\theta_0|\leq\Delta} E_{\underset{\sim}{\theta}}[\delta_B(\underset{\sim}{X})] = \lim_{|\mu|\to\infty} g(\mu) = 0.$$ Also, when

$\beta_1^2 \neq \beta_0^2$, $\{\underset{\sim}{\theta}|\mu=0\} \cap \{\underset{\sim}{\theta}||\theta_1-\theta_0| \geq \Delta+\varepsilon\} \neq \phi$.

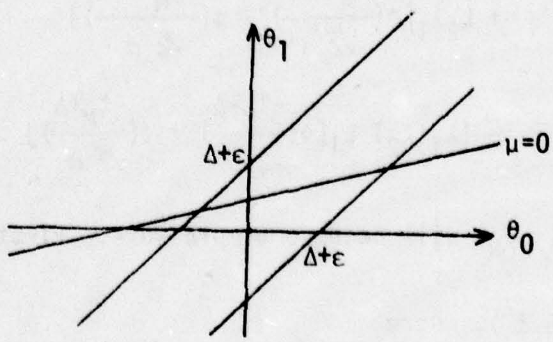


Figure 4. Graph of $\mu = 0$ on $\theta_0\theta_1$-plane.

Hence,

$$\sup_{|\theta_1-\theta_0|\geq\Delta+\varepsilon} E_{\underset{\sim}{\theta}}[\delta_B(\underset{\sim}{X})] = g(0) = \Phi(\frac{t_B}{\zeta}) - \Phi(\frac{-t_B}{\zeta}) .$$

By Lemma 1.8.1.,

$$\sup_{\tau\in\Gamma} r^{(1)}(\tau,\delta_B) = L_1\lambda_1 + L_2\lambda_1[\Phi(\frac{t_B}{\zeta}) - \Phi(\frac{-t_B}{\zeta})] . \qquad (1.8.2)$$

(b) If $\beta_1^2 = \beta_0^2 = \beta^2$, then $\mu = \frac{\beta^2}{\beta^2 + \sigma^2}(\theta_1-\theta_0) + \frac{\sigma^2(\alpha_1-\alpha_0)}{\beta^2 + \sigma^2}$.

(i) If $\alpha_1 > \alpha_0$, then under $|\theta_1-\theta_0| \leq \Delta$, $|\mu|$ has its largest value when $\theta_1 - \theta_0 = \Delta$. In this case, let $\mu_1 = \frac{\beta^2\Delta+\sigma^2(\alpha_1-\alpha_0)}{\beta^2 + \sigma^2}$, so

$$\inf_{|\theta_1-\theta_0| \leq \Delta} E_{\underset{\sim}{\theta}}[\delta_B(\underset{\sim}{X})] = \Phi(\frac{t_0-\mu_1}{\zeta}) - \Phi(\frac{-t_B-\mu_1}{\zeta}) .$$

If $\alpha_1 \leq \alpha_0$, under $|\theta_1-\theta_0| \leq \Delta$, $|\mu|$ has its largest value if $\theta_1-\theta_0 = -\Delta$, then

$$\inf_{|\theta_1-\theta_0| \leq \Delta} E_{\underset{\sim}{\theta}}[\delta_B(\underset{\sim}{X})] = \Phi(\frac{t_B-\mu_2}{\zeta}) - \Phi(\frac{-t_B-\mu_2}{\zeta})$$

where

$$\mu_2 = \frac{\beta^2\Delta+\sigma^2(\alpha_0-\alpha_1)}{\beta^2 + \sigma^2} .$$

So we get

$$\inf_{|\theta_1-\theta_0| \leq \Delta} E_{\underset{\sim}{\theta}}[\delta_B(\underset{\sim}{X})] = \Phi(\frac{t_B-\mu_0}{\zeta}) - \Phi(\frac{-t_B-\mu_0}{\zeta}) ,$$

where

$$\mu_0 = \frac{\Delta\beta^2+\sigma^2|\alpha_1-\alpha_0|}{\beta^2 + \sigma^2} .$$

(ii) To find $\sup_{|\theta_1-\theta_0| \geq \Delta+\varepsilon} E_{\underset{\sim}{\theta}}[\delta_B(\underset{\sim}{X})]$, let us note that $\mu=0$ iff $\frac{\sigma^2}{\beta^2}(\alpha_0-\alpha_1) = \theta_1-\theta_0$. Then

(1) If $\frac{\sigma^2}{\beta^2}|\alpha_0-\alpha_1| \geq \Delta+\varepsilon$, then $\{\mu=0\} \cap \{|\theta_1-\theta_0| \geq \Delta+\varepsilon\} \neq \phi$.

So, $\underset{|\theta_1-\theta_0| \geq \Delta+\varepsilon}{} E_{\underset{\sim}{\theta}}[\delta_B(\underset{\sim}{X})] = \Phi(\frac{t_B}{\zeta}) - \Phi(\frac{-t_B}{\zeta})$.

(2) If $\frac{\sigma^2}{\beta^2}|\alpha_0-\alpha_1|<\Delta+\varepsilon$, then under $|\theta_1-\theta_0|\geq\Delta+\varepsilon$, $|\mu|$ has the smallest value when $\mu=\mu' = \dfrac{\beta^2(-\Delta-\varepsilon)+\sigma^2|\alpha_1-\alpha_0|}{\beta^2+\sigma^2}$, then

$$\sup_{|\theta_1-\theta_0|\geq\Delta+\varepsilon} E_\theta[\delta_B(\underset{\sim}{X})] = \Phi\left(\frac{t_B-\mu_1}{\zeta}\right) - \Phi\left(\frac{-t_B-\mu_1}{\zeta}\right).$$

To sum up, if $\beta_1^2 = \beta_0^2 = \beta^2$, then

$$\sup_{\tau\in\Gamma} r^{(1)}(\tau,\delta_B) = \begin{cases} L_1\lambda_1\left[\Phi\left(\frac{-t_B+\mu_0}{\zeta}\right) + \left(\frac{-t_B-\mu_0}{\zeta}\right)\right] + L_2\left[\Phi\left(\frac{t_B}{\zeta}\right) - \Phi\left(\frac{-t_B}{\zeta}\right)\right] \\ \qquad \text{if } \frac{\sigma^2}{\beta^2}|\alpha_0-\alpha_1|\geq\Delta+\varepsilon \\ L_1\lambda_1\left[\Phi\left(\frac{-t_B+\mu_0}{\zeta}\right) + \Phi\left(\frac{-t_B-\mu_0}{\zeta}\right)\right] + L_2\left[\Phi\left(\frac{t_B-\mu'}{\zeta}\right) - \Phi\left(\frac{-t_B-\mu'}{\zeta}\right)\right] \\ \qquad \text{if } \frac{\sigma^2}{\beta^2}|\alpha_0-\alpha_1|<\Delta+\varepsilon \end{cases}$$

At last, let $\Theta^* = \{\tau|\tau$ is a distribution on $\Theta\}$, we want to compute $\sup_{\tau\in\Theta^*} r^{(1)}(\tau,\delta_B)$, $\sup_{\tau\in\Theta^*} r^{(1)}(\tau,\delta_G)$ and $\sup_{\tau\in\Theta^*} r^{(1)}(\tau,\delta_M)$.

Lemma 1.8.2.

$$\sup_{\tau\in\Theta^*} r^{(1)}(\tau,\delta) = \max\left[L_1\left(1-\inf_{|\theta_1-\theta_0|\leq\Delta} E_\theta[\delta(\underset{\sim}{X})]\right), L_2 \sup_{|\theta_1-\theta_0|\geq\Delta+\varepsilon} E_\theta[\delta(\underset{\sim}{X})]\right]$$

Proof: From (1.8.1), for all $\tau\in\Theta^*$

$$r^{(1)}(\tau,\delta) \leq L_1\left(1-\inf_{|\theta_1-\theta_0|\leq\Delta} E_\theta[\delta(\underset{\sim}{X})]\right)P_\tau[|\theta_1-\theta_0|\leq\Delta]$$

$$+ L_2 \sup_{|\theta_1-\theta_0|\geq\Delta+\varepsilon} E_\theta[\delta(\underset{\sim}{X})]P_\tau[|\theta_1-\theta_0|\geq\Delta+\varepsilon]$$

$$\leq \max \left(L_1\left(1- \inf_{|\theta_1-\theta_0|\leq\Delta} E_{\theta}[\delta(\underset{\sim}{X})]\right),\ L_2 \sup_{|\theta_1-\theta_0|\geq\Delta+\varepsilon} E_{\theta}[\delta(\underset{\sim}{X})]\right).$$

Now, let $\underset{\sim}{\theta}_n$ and $\underset{\sim}{\theta}_n'$ be the same as in Lemma 1.8.1, but we let τ_n be such that

$$P_{\tau_n}[\underset{\sim}{\theta}=\underset{\sim}{\theta}_n] = 1 \quad \text{if} \quad L_1(1-E_{\underset{\sim}{\theta}_n}[\delta(\underset{\sim}{X})]) \geq L_2E_{\underset{\sim}{\theta}_n'}[\delta(\underset{\sim}{X})]$$

and

$$P_{\tau_n}[\underset{\sim}{\theta}=\underset{\sim}{\theta}_n'] = 1 \quad \text{if} \quad L_1(1-E_{\underset{\sim}{\theta}_n}[\delta(\underset{\sim}{X})] \leq L_2E_{\underset{\sim}{\theta}_n'}[\delta(\underset{\sim}{X})] ,$$

then

$$r^{(1)}(\tau_n,\delta) = \max(L_1(1-E_{\underset{\sim}{\theta}_n}[\delta(\underset{\sim}{X})]),\ L_2E_{\underset{\sim}{\theta}_n'}[\delta(\underset{\sim}{X})])$$

$$\rightarrow \max(L_1(1- \inf_{|\theta_1-\theta_0|\leq\Delta} E_{\theta}[\delta(\underset{\sim}{X})],\ L_2 \sup_{|\theta_1-\theta_0|\geq\Delta+\varepsilon} E_{\theta}[\delta(\underset{\sim}{X})]) .$$

This finishes the proof.

From Theorem 1.7.2, we get

$$\sup_{\theta\in\Theta^*} r^{(1)}(\tau,\delta_M) = L_1[\Phi(\frac{-t_M-\Delta}{\sqrt{2}\,\sigma}) + \Phi(\frac{-t_M+\Delta}{\sqrt{2}\,\sigma})]$$

$$= L_2[\Phi(\frac{t_M-\Delta-\varepsilon}{\sqrt{2}\,\sigma}) - \Phi(\frac{-t_M-\Delta-\varepsilon}{\sqrt{2}\,\sigma})] .$$

From Lemma 1.8.2., it is obvious that

$$\sup_{\theta\in\Theta^*} r^{(1)}(\tau,\delta_G) = \max(L_1[\Phi(\frac{-t_G-\Delta}{\sqrt{2}\,\sigma}) + \Phi(\frac{-t_G+\Delta}{\sqrt{2}\,\sigma})],$$

$$L_2[\Phi(\frac{t_G-\Delta-\varepsilon}{\sqrt{2}\,\sigma}) - \Phi(\frac{-t_G-\Delta-\varepsilon}{\sqrt{2}\,\sigma})]).$$

And if $\beta_0^2 \neq \beta_1^2$, then

$$\sup_{\tau^*} r(\tau,\delta_B)=\max(L_1,\ L_2[\Phi(\frac{t_B}{\zeta}) - \Phi(\frac{-t_B}{\zeta})) . \qquad (1.8.3)$$

If $\beta_0^2 = \beta_1^2 = \beta^2$, then

$$\sup_{\tau\in\Theta^*} r(\tau,\delta_B) = \begin{cases} \max(L_1[\Phi(\frac{-t_B+\mu_0}{\zeta}) + \Phi(\frac{-t_B-\mu_0}{\zeta})],\ L_2[\Phi(\frac{t_B}{\zeta}) - \Phi(\frac{-t_B}{\zeta})]) \\ \qquad \text{if } \frac{\sigma^2}{\beta^2}|\alpha_0-\alpha_1| \geq \Delta+\varepsilon \\ \max(L_1[\Phi(\frac{-t_B+\mu_0}{\zeta}) + \Phi(\frac{-t_B-\mu_0}{\zeta})],\ L_2[\Phi(\frac{t_B-\mu'}{\zeta}) - \Phi(\frac{-t_B-\mu'}{\zeta})]) \\ \qquad \text{if } \frac{\sigma^2}{\beta^2}|\alpha_0-\alpha_1| < \Delta+\varepsilon . \end{cases}$$

Remark: All the risk computed in this section are based on one sample from each population. If we have n samples from each population, by reducing to sufficient statistic, we only need to change σ^2 to $\frac{\sigma^2}{n}$, and all the formulas will remain valid. The formulas are used to compute the following tables.

Illustration of the table:

(1) The control parameter θ_0 is assumed to have prior distribution as $N(0,1)$, and θ_1 is assumed to be distributed as $N(\alpha,\beta^2)$, where (α,β^2) are chosen as (1,1), (0,.5), (0,1) and (0,2) in the tables.

(2) $\frac{\sigma^2}{n}$ are chosen as .2 in Table I and as .5 in Table II.

(3) Δ are chosen as .5, 1., and 1.5.

(4) For Δ=.5, ε are chosen as .2 and .4.
For Δ=1., ε are chosen as .3 and .8.
For Δ=1.5, ε are chosen as .5 and 1.

(5) When (α,β^2), $\frac{\sigma^2}{n}$, Δ and ε are fixed, λ and λ' are computed so that $\tau_B \in \Gamma$. t_B, t_G and t_M are found, and $r(\tau_B,\delta)$, $\sup_{\tau\in\Gamma} r(\tau,\delta)$, and $\sup_{\tau\in\Theta^*} r(\tau,\delta)$ for $\delta = \delta_B$, δ_G, δ_M are computed.

They are arranged in the following manner:

t_B	$r(\tau_B,\delta_B)$	$\sup_{\tau\in\Gamma} r(\tau,\delta_B)$	$\sup_{\tau\in\Theta^*} r(\tau,\delta_B)$	1.0	$\dfrac{\sup_{\tau\in\Gamma} r(\tau,\delta_B)}{\sup_{\tau\in\Gamma} r(\tau,\delta_G)}$	$\dfrac{\sup_{\tau\in\Theta^*} r(\tau,\delta_B)}{\sup_{\tau\in\Theta^*} r(\tau,\delta_M)}$
t_G	$r(\tau_B,\delta_G)$	$\sup_{\tau\in\Gamma} r(\tau,\delta_G)$	$\sup_{\tau\in\Theta^*} r(\tau,\delta_G)$	$\dfrac{r(\tau_B,\delta_G)}{r(\tau_B,\delta_B)}$	1.0	$\dfrac{\sup_{\tau\in\Theta^*} t(\tau,\delta_G)}{\sup_{\tau\in\Theta^*} r(\tau,\delta_M)}$
t_M	$r(\tau_B,\delta_M)$	$\sup_{\tau\in\Gamma} r(\tau,\delta_M)$	$\sup_{\tau\in\Theta^*} r(\tau,\delta_M)$	$\dfrac{r(\tau_B,\delta_M)}{r(\tau_B,\delta_B)}$	$\dfrac{\sup_{\tau\in\Gamma} r(\tau,\delta_M)}{\sup_{\tau\in\Gamma} r(\tau,\delta_G)}$	1.0

(6) All tables are computed under the assumption that $L_1 = L_2$.

To use the table:

(a) For the Bayes rule: For specified values of $\frac{\sigma^2}{n}$, (α,β^2), Δ and ε, look for t_B and the risks in the first row of each block.

(b) For the Γ-minimax rule: For specified values of $\frac{\sigma^2}{n}$, Δ, ε, λ and λ', look for t_G and the risks in the second row of each block.

(c) For the minimax rule: For specified values of $\frac{\sigma^2}{n}$, Δ and ε , look for the t_M and the risks in the third row of each block.

Table I.1 1. The first column lists values of t_B, t_G, and t_M.

2. The second block of numbers are values of $r(\tau_B,\delta)$, $\sup_{\tau\in\Gamma} r(\tau,\delta)$ and $\sup_{\tau\in\Gamma^*} r(\tau,\delta)$ corresponding to $\delta=\delta_B,\delta_G$ and δ_M.

3. The entries in the third block are values of ratios of the risks in the second block (dividing each column by the diagonal value).

$\sigma^2/n = .2$, $(\alpha, \beta^2) = (1,1)$

$\Delta = .5$		$\varepsilon = .2,\ \lambda = .2174$			$\lambda' = .6987$		
	.5657	.1529	.5201	.5802		2.3924	1.2991
	0.	.2174	.2174	1.0	1.4216		2.2391
	.6421	.1598	.4091	.4466	1.0450	1.8818	
		$\varepsilon = .4,\ \lambda = .2174$			$\lambda' = .6177$		
	.6820	.1193	.4440	.5661		2.0424	1.4642
	0.	.2174	.2174	1.0	1.8224		2.5865
	.7261	.1260	.3229	.3866	1.0562	1.4851	
$\Delta = 1.0$		$\varepsilon = .3,\ \lambda = .4214$			$\lambda' = .4679$		
	1.1499	.1224	.4775	.6709		1.3280	1.6509
	1.0166	.1420	.3595	.4902	1.1593		1.2063
	1.1503	.1312	.3614	.4064	1.0711	1.0052	
		$\varepsilon = .8,\ \lambda = .4214$			$\lambda' = .3097$		
	1.4000	.0575	.2648	.5503		1.4065	2.0878
	1.5542	.0609	.1883	.3488	1.0593		1.3232
	1.4001	.0649	.1927	.2636	1.1297	1.0235	
$\Delta = 1.5$		$\varepsilon = .5,\ \lambda = .5996$			$\lambda' = .2567$		
	1.7500	.0735	.3332	.6824		1.4156	1.9704
	2.4287	.0999	.2354	.7511	1.3594		2.1687
	1.7500	.0827	.2966	.3463	1.1253	1.2600	
		$\varepsilon = 1.0,\ \lambda = .5996$			$\lambda' = .1511$		
	2.0000	.0308	.1655	.5628		1.5135	2.6227
	2.5514	.0390	.1093	.5324	1.2652		2.4808
	2.0000	.0383	.1611	.2146	1.2445	1.4733	

Table I (Continued), Table I.2

$\sigma^2/n = .2,\ (\alpha,\beta^2) = (0,\ .5)$

		$\varepsilon = .2,\ \lambda = .3169$			$\lambda' = .5676$		
	.5731	.2056	.7466	1.0		2.3559	2.2391
	0.	.3169	.3169	1.0	1.5414		2.2391
	.6421	.2097	.3950	.4466	1.0199	1.2465	
$\Delta = .5$		$\varepsilon = .4,\ \lambda = .3169$			$\lambda' = .4624$		
	.6861	.1539	.7043	1.0		2.4115	2.5865
	.4766	.1931	.2921	.5760	1.2540		1.4899
	.7261	.1611	.3013	.3866	1.0464	1.0316	
		$\varepsilon = .3,\ \lambda = .5858$			$\lambda' = .2885$		
	1.1500	.1303	.8687	1.0		3.0719	2.4606
	2.0944	.1798	.2828	.8955	1.3802		2.2034
	1.1503	.1512	.3553	.4064	1.1608	1.2564	
$\Delta = 1.$		$\varepsilon = .8,\ \lambda = .5858$			$\lambda' = .1416$		
	1.4000	.0508	.7268	1.0		6.0236	3.7939
	2.1098	.0619	.1207	.6879	1.2167		2.6098
	1.4001	.0708	.1917	.2636	1.3920	1.5891	
		$\varepsilon = .5,\ \lambda = .7793$			$\lambda' = .1025$		
	1.7500	.0524	.8818	1.0		8.6338	2.8875
	3.3731	.0889	.1021	.9850	1.6986		2.8443
	1.7500	.0771	.3054	.3463	1.4732	2.9901	
$\Delta = 1.5$		$\varepsilon = 1.0,\ \lambda = .7793$			$\lambda' = .0412$		
	2.0000	.0169	.8206	1.0		21.3210	4.6599
	3.1757	.0264	.0385	.8573	1.5578		3.9951
	2.0000	.0351	.1761	.2146	2.0717	4.5754	

Table I (Continued), Table I.3

$\sigma^2/n = .2, \ (\alpha, \beta^2) = (0, 1)$

$\Delta = .5$		$\epsilon = .2, \lambda = .2763$			$\lambda' = .6206$		
	.5657	.1902	.4090	.4721		1.4800	1.0570
	0.	.2763	.2763	1.0	1.4525		2.2391
	.6421	.1905	.4006	.4466	1.0016	1.4496	
		$\epsilon = .4, \lambda = .2763$			$\lambda' = .5245$		
	.6820	.1459	.3237	.4454		1.1791	1.1520
	.1873	.2297	.2745	.8281	1.5743		2.1419
	.7261	.1479	.3096	.3866	1.0132	1.1279	
$\Delta = 1.$		$\epsilon = .3, \lambda = .5205$			$\lambda' = .3580$		
	1.1499	.1337	.3397	.5503		1.0190	1.3540
	1.6494	.1461	.3333	.7097	1.0928		1.7463
	1.1503	.1447	.3570	.4064	1.0818	1.0711	
		$\epsilon = .8, \lambda = .5205$			$\lambda' = .2031$		
	1.4000	.0576	.1597	.4247		1.0341	1.6114
	1.8706	.0619	.1545	.5444	1.0743		2.0656
	1.4001	.0690	.1907	.2636	1.1979	1.2347	
$\Delta = 1.5$		$\epsilon = .5, \lambda = .7112$			$\lambda' = .1573$		
	1.7500	.0647	.2104	.5628		1.3610	1.6252
	2.9570	.1066	.1546	.9349	1.6478		2.6995
	1.7500	.0804	.3008	.3463	1.2433	1.9453	
		$\epsilon = 1.0, \lambda = .7112$			$\lambda' = .0771$		
	2.0	.0241	.0887	.4372		1.3378	2.0372
	2.8887	.0344	.0663	.7306	1.4276		3.4045
	2.0	.0367	.1692	.2146	1.5211	2.5505	

Table I (Continued), Table I.4

$\sigma^2/n = .2$, $(\alpha,\beta^2) = (0, 2)$

		$\varepsilon = .2$, $\lambda = .2272$			$\lambda' = .6861$		
	.5605	.1627	.7009	1.0		3.0855	2.2391
	0.	.2272	.2272	1.0	1.3966		2.2391
	.6421	.1638	.4079	.4466	1.0072	1.7954	
$\Delta = .5$		$\varepsilon = .4$, $\lambda = .2272$			$\lambda' = .6033$		
	.6791	.1271	.6989	1.0		3.0764	2.5865
	0.	.2272	.2272	1.0	1.7876		2.5865
	.7261	.1286	.3211	.3866	1.0120	1.4134	
		$\varepsilon = .3$, $\lambda = .4363$			$\lambda' = .4529$		
	1.1499	.1260	.8724	1.0		2.4156	2.4606
	1.1044	.1350	.3612	.4349	1.0719		1.0700
	1.1503	.1318	.3614	.4064	1.0459	1.0006	
$\Delta = 1.$		$\varepsilon = .8$, $\lambda = .4363$			$\lambda' = .2986$		
	1.4000	.0588	.7317	1.0		3.9116	3.7939
	1.5896	.0597	.1871	.3697	1.0138		1.4026
	1.4001	.0648	.1937	.2636	1.1017	1.0357	
		$\varepsilon = .5$, $\lambda = .6135$			$\lambda' = .2482$		
	1.7500	.0725	.8614	1.0		3.7480	2.8875
	2.4739	.0982	.2298	.7732	1.3543		2.2326
	1.7500	.0810	.2984	.3463	1.1174	1.2985	
$\Delta = 1.5$		$\varepsilon = 1.0$, $\lambda = .6135$			$\lambda' = .1489$		
	2.0	.0304	.7624	1.0		7.0051	4.6599
	2.5663	.0375	.1088	.5418	1.2320		2.5246
	2.0	.0376	.1636	.2146	1.2364	1.5034	

Table II.1 1. The first column list values of t_B, t_G and t_M.

2. The second block of numbers are values of $r(\tau_B,\delta)$, $\sup_{\tau\,\Gamma} r(\tau,\delta)$ and $\sup_{\tau\,*} r(\tau,\delta)$ corresponding to $\delta=\delta_B,\delta_G$ and δ_M.

3. The entries in the third block are values of ratios of the risks in the second block (dividing each column by the diagonal value).

$\sigma^2/n = .5$, $(\alpha, \beta^2) = (1, 1)$

$\Delta = .5$		$\varepsilon = .2$, $\lambda = .2174$			$\lambda' = .6987$		
	.3275	.2104	.4242	.7625		1.9511	1.6022
	0.	.2174	.2174	1.0	1.0334		2.1013
	.8054	.2382	.4360	.4759	1.1321	2.0051	
		$\varepsilon = .4$, $\lambda = .2174$			$\lambda' = .6177$		
	.5571	.1848	.4761	.5985		2.1900	1.3432
	0.	.2174	.2174	1.0	1.1767		2.2444
	.8615	.2038	.3721	.4456	1.1030	1.7115	
$\Delta = 1.$		$\varepsilon = .3$, $\lambda = .4214$			$\lambda' = .4679$		
	1.1441	.1951	.5559	.8143		1.4109	1.8320
	.9872	.2358	.3940	.5286	1.2084		1.1892
	1.1771	.2199	.3953	.4445	1.1270	1.0031	
		$\varepsilon = .8$, $\lambda = .4214$			$\lambda' = .3097$		
	1.3980	.1148	.3597	.7869		1.4676	2.2599
	1.8158	.1258	.2451	.5062	1.0958		1.4535
	1.4117	.1404	.2546	.3482	1.2230	1.0385	
$\Delta = 1.5$		$\varepsilon = .5$, $\lambda = .5996$			$\lambda' = .2567$		
	1.7500	.1264	.3827	.8697		1.5118	2.1659
	3.4468	.1864	.2532	.9260	1.4749		2.3062
	1.7509	.1578	.3438	.4015	1.2491	1.3581	
		$\varepsilon = 1.0$, $\lambda = .5996$			$\lambda' = .1511$		
	2.0000	.0674	.2222	.8413		1.5820	2.7259
	3.3785	.0902	.1405	.8102	1.3382		2.6248
	2.0003	.0972	.2317	.3087	1.4427	1.6493	

Table II (Continued), Table II.2

$\sigma^2/n = .5, \quad (\alpha,\beta^2) = (0, .5)$

$\Delta = .5$		$\varepsilon = .2,\ \lambda = .3169$			$\lambda' = .5676$		
	.4188	.2880	.6136	1.0		1.9363	2.1014
	0.	.3169	.3169	1.0	1.1002		2.1014
	.8054	.2930	.4209	.4759	1.0171	1.3283	
		$\varepsilon = .4,\ \lambda = .3169$			$\lambda' = .4624$		
	.6014	.2386	.6372	1.0		2.0105	2.2444
	0.	.3169	.3169	1.0	1.3281		2.2444
	.8615	.2443	.3472	.4456	1.0236	1.0957	
$\Delta = 1.$		$\varepsilon = .3,\ \lambda = .5858$			$\lambda' = .2885$		
	1.1469	.1968	.8594	1.0		2.9828	2.2499
	3.5136	.2654	.2881	.9866	1.3485		2.2197
	1.1771	.2522	.3886	.4447	1.2814	1.3487	
		$\varepsilon = .8,\ \lambda = .5858$			$\lambda' = .1416$		
	1.3991	.0928	.7249	1.0		5.2358	2.8717
	3.1767	.1131	.1384	.9157	1.2183		2.6296
	1.4117	.1562	.2533	.3482	1.6834	1.8309	
$\Delta = 1.5$		$\varepsilon = .5,\ \lambda = .7793$			$\lambda' = .4990$		
	1.7500	.0803	.8815	1.0		8.6025	2.4904
	5.8077	.1022	.1025	.9999	1.2727		2.4902
	1.7509	.1579	.3541	.4015	1.9662	3.4554	
		$\varepsilon = 1.,\ \lambda = .7793$			$\lambda' = .0412$		
	2.0000	.0313	.8205	1.0		19.9395	3.2399
	4.9393	.0398	.0412	.9926	1.2711		3.2161
	2.0003	.0979	.2533	.3087	3.1252	6.1546	

Table II (Continued), Table II.3

$\sigma^2/n = .5$, $(\alpha,\beta^2) = (0, 1)$

	$\varepsilon = .2$, $\lambda = .2763$				$\lambda' = .6206$		
	.3275	.2661	.3701	.6643		1.3393	1.3958
	0.	.2763	.2763	1.0	1.0382		2.1014
	.8054	.2728	.4268	.4759	1.0248	1.5447	
$\Delta = .5$	$\varepsilon = .4$, $\lambda = .2763$				$\lambda' = .5245$		
	.5571	.2296	.3541	.4594		1.2814	1.0311
	0.	.2763	.2763	1.0	1.2034		2.2444
	.8615	.2297	.3568	.4456	1.0002	1.2913	
	$\varepsilon = .3$, $\lambda = .5205$				$\lambda' = .3580$		
	1.1441	.2113	.3613	.6601		1.0278	1.4851
	2.4180	.2430	.3516	.8681	1.1500		1.9532
	1.1771	.2412	.3904	.4445	1.1416	1.1106	
$\Delta = 1.$	$\varepsilon = .8$, $\lambda = .5205$				$\lambda' = .2031$		
	1.3980	.1121	.1967	.6167		1.0427	1.7711
	2.5834	.1238	.1887	.7833	1.1044		2.2494
	1.4117	.1506	.2520	.3482	1.3437	1.3356	
	$\varepsilon = .5$, $\lambda = .7112$				$\lambda' = .1573$		
	1.7500	.1066	.2081	.7340		1.3237	1.8280
	4.7675	.1517	.1572	.9972	1.4228		2.4834
	1.7509	.1592	.3487	.4015	1.4929	2.2177	
$\Delta = 1.5$	$\varepsilon = 1.0$, $\lambda = .7112$				$\lambda' = .0771$		
	2.0000	.0493	.1008	.6915		1.3244	2.2402
	4.2218	.0659	.0761	.9574	1.3379		3.1020
	2.0003	.0979	.2433	.3087	1.9879	3.1959	

Table II (Continued), Table II.4

$\sigma^2/n = .5, \quad (\alpha, \beta^2) = (0, 2)$

		$\varepsilon = .2,\ \lambda = .2272$			$\lambda' = .6861$		
$\Delta = .5$	.2161	.2251	.3855	1.0		1.6970	2.1014
	0.	.2272	.2272	1.0	1.0091		2.1014
	.8054	.2417	.4346	.4759	1.0739	1.9131	
		$\varepsilon = .4,\ \lambda = .2272$			$\lambda' = .6033$		
	.5137	.2011	.5377	1.0		2.3667	2.2444
	0.	.2272	.2272	1.0	1.1298		2.2444
	.8615	.2061	.3700	.4456	1.0249	1.6289	
$\Delta = 1.$		$\varepsilon = .3,\ \lambda = .4363$			$\lambda' = .4529$		
	1.1409	.2058	.8343	1.0		2.1109	2.2499
	1.1754	.2212	.3952	.4452	1.0744		1.0016
	1.1771	.2210	.3952	.4445	1.0739	1.0000	
		$\varepsilon = .8,\ \lambda = .4363$			$\lambda' = .4895$		
	1.3968	.1205	.7177	1.0		2.9617	2.8717
	1.8999	.1230	.2423	.5397	1.0205		1.5498
	1.4117	.1409	.2559	.3482	1.1686	1.0562	
$\Delta = 1.5$		$\varepsilon = .5,\ \lambda = .6135$			$\lambda' = .2482$		
	1.7459	.1269	.8574	1.0		3.4916	2.4904
	3.5599	.1833	.2456	.9406	1.4440		2.3425
	1.7509	.1563	.3460	.4015	1.2312	1.4091	
		$\varepsilon = 1,\ \lambda = .6135$			$\lambda' = .1489$		
	2.0000	.0679	.7615	1.0		5.4734	3.2399
	3.4159	.0878	.1391	.8201	1.2924		2.6572
	2.0003	.0967	.2353	.3087	1.4249	1.6915	

1.9 An example and conclusion

After we study the tables computed in Section 1.8, some trends can be found:

1. If $\binom{\alpha}{\beta^2} = \binom{0}{1}$, then the Bayes rule performs very well in terms of $\sup_{\tau\in\Gamma} r(\tau,\delta_B)$ and $\sup_{\tau\in\Theta^*} r(\tau,\delta_B)$; this is because when $\binom{\alpha}{\beta^2} = \binom{0}{1}$, $|a_1-a_0| = \frac{1}{1+\sigma^2}|x_1-x_0|$, hence Bayes rule has the same form as Γ-minimax rule and the minimax rule.

2. If $\beta^2 \neq 1$, the Bayes rule does not perform as well. This was shown in formula (1.8.2) and (1.8.3). One can also find that when $\frac{\sigma^2}{n} = .5$, $\binom{\alpha}{\beta^2} = \binom{0}{.5}$, $\Delta = 1.5$, and $\varepsilon = 1$,

$$\frac{\sup_{\tau\in\Gamma} r(\tau,\delta_B)}{\sup_{\tau\in\Gamma} r(\tau,\delta_G)} = 19.9395 .$$

This means a large increase in loss will occur if we need to consider $\sup_{\tau\in\Gamma}(\tau,\delta_B)$ instead of $r(\tau_B,\delta_B)$.

3. Γ-minimax rule is robust in terms of $\sup_{\tau\in\Theta^*} r(\tau,\delta_G)$ if λ_1 and λ_1' are close to each other. This is because $\sup_{\tau\in\Theta^*} r(\tau,\delta_G) = \max(A,B)$, but $\sup_{\tau\in\Gamma} r(\tau,\delta_G) = \lambda_1 A + \lambda_1' B$. Also from the tables, for all λ and λ', $r(\tau_B,\delta_G) \leq 2\, r(\tau_B,\delta_B)$.

4. Minimax rule in general performs fairly well.

5. Γ-minimax rule is not necessarily better than the minimax rule in terms of the Bayes risk. This means that when the decision depends on full information, sometimes incomplete information is worse than no information.

6. When ε gets larger, all risks become smaller.

Example 1.9.1: A company has type Π_0 machines to produce part P(p) [p is the diameter of P] and $p|\Pi_0 \sim N(\theta_0 \times 10^{-2}$ in., 1×10^{-4} sq. in.). However, the same company also has type Π_1, Π_2, Π_3 machines which produce part Q(q) and $q|\Pi_i \sim N(\theta_i \times 10^{-2}$ in., 1×10^{-4} sq. in.). P and Q are matched if $|p-q| \leq .045$ in. Since $|\theta_i - \theta_0| \leq 1.5 > P[|p-q| \leq 4.5] = P\left[\frac{-4.5-(\theta_i-\theta_0)}{\sqrt{2}} \leq Z \leq \frac{4.5-(\theta_i-\theta_0)}{\sqrt{2}}\right] \geq P\left[\frac{-6}{\sqrt{2}} \leq Z \leq \frac{3}{\sqrt{2}}\right] = .98$, so we can define Π_i as good for Π_0 iff $|\theta_i - \theta_0| \leq 1.5$. Similarly, we would like to define Π_i as bad for Π_0 iff $|\theta_i = \theta_0| \geq 2.5$. The company claims:

$P[|\theta_1 - \theta_0| \leq 1.5] = .78$, $P[|\theta_1 - \theta_0| \geq 2.5] = .04$

$P[|\theta_2 - \theta_0| \leq 1.5] = .71$, $P[|\theta_2 - \theta_0| \geq 2.5] = .08$

$P[|\theta_3 - \theta_0| \leq 1.5] = .61$, $P[|\theta_3 - \theta_0| \geq 2.5] = .15$.

Now, the company has machines a_0, a_1, a_2, a_3 for sale, where $a_i \in \Pi_i$, for $i = 0,1,2,3$. If we are allowed to take 5 sample parts from each machine, which machines to produce part Q should we buy?

Solution: Let $\overline{X}_0$, $\overline{X}_1$, $\overline{X}_2$, $\overline{X}_3$ be the mean observation from a_0, a_1, a_2, a_3, respectively. Then $\frac{\sigma^2}{n}$, $\Delta = 1.5$, $\varepsilon = 1.0$, and if we decide to use Γ-minimax rule, the table I.2, I.3 and I.4 indicate

a_1 is good for a_0 iff $|\overline{X}_1 - \overline{X}_0| \leq 3.1757$

a_2 is good for a_0 iff $|\overline{X}_2 - \overline{X}_0| \leq 2.8887$

and a_3 is good for a_0 iff $|\overline{X}_3 - \overline{X}_0| \leq 2.5663$.

If we feel the claims made by the company may not be correct and we would rather assume there is no prior information, then we will decide to use minimax decision rule. Then

a_i is good for a_0 iff $|\overline{X}_i-\overline{X}_0| \le 2.0$ for all i=1,2,3.

If from another source, we know more informations such that $\theta_0 \sim N(c, 1)$, $\theta_1 \sim N(c, .5)$, $\theta_2 \sim N(c, 1)$, and $\theta_3 \sim N(c, 2)$, for some c, then we would like to use the Bayes rule. So

$$a_1 \text{ is good for } a_0 \text{ iff } \left|\frac{c\cdot\frac{1}{5} + \overline{X}_1(.5)}{\frac{1}{5} + .5} - \frac{c\cdot\frac{1}{5} + \overline{X}_0(1)}{\frac{1}{5} + 1}\right| \le 2.0$$

$$\text{i.e. } \left|\frac{2c + 5\,\overline{X}_1}{7} - \frac{c + 5\,\overline{X}_0}{6}\right| \le 2.0$$

a_2 is good for a_0 iff $|\overline{X}_2 - \overline{X}_0| \le 2.40$

and a_3 is good for a_0 iff $\left|\frac{c + 10\,\overline{X}_2}{11} - \frac{c + 5\,\overline{X}_0}{6}\right| \le 2.0.$

If we suspect the definiteness of any prior information, we may then use the rule which is most robust to the assumption of the prior distribution. So from the table we use Γ-minimax rule on a_1, use Bayes rule on a_2, and use the minimax rule on a_3.

CHAPTER II

Γ-MINIMAX RULES FOR SELECTING THE t-BEST POPULATIONS

2.1 Introduction

In this chapter, we continue further investigations of the Γ-minimax procedures. The problem considered here is to select the t-best populations out of k populations for some fixed $t < k$. Deverman and Gupta (1969), Carroll, Gupta and Huang (1975) have discussed this problem under the subset selection approach. Carroll and Gupta (1977) also provided an algorithm which can be used to compute the ranking probability $P\{(X_1,\dots,X_{t_1}) < (X_{t_1+1},\dots,X_{t_2}) < \dots < (X_{t_s+1},\dots,X_k)\}$, where X_i has pdf $f(x-\theta_i)$ and $\theta_1 = \dots = \theta_{t_1} < \theta_{t_1+1} = \dots = \theta_{t_2} < \dots < \theta_{t_s+1} = \dots = \theta_k$. For the problem of selecting exactly t population, Bahadur and Goodman (1952) and Alam (1973) have shown some optimal properties of the natural selection procedure.

In Section 2.2, it is shown that if the populations have PF_2 densities, then the natural selection rule is a Γ-minimax rule. This result is also extended to the case when the populations are not required to be independent but have some particular form. This is done in Section 2.4. In Section 2.3, our goal is to rank the k populations through a simultaneous selection of the t-best populations for all $1 \le t \le k - 1$. In order that a Γ-minimax rule can be obtained, we need to change the loss function we used in Section 2.2 slightly, so that an indifference zone is allowed.

The result obtained in this section justifies why we adopt midrank for tied data. In the last section of this chapter, the result of Gupta and Huang (1977) is generalized and it is shown that the Γ-minimax rule for selecting the best population can be found even if the populations are not independent. We also prove a lemma which will help us to find the Γ-minimax rules for testing hypotheses about multinomial distributions and multivariate negative binomial distributions.

2.2 Selecting the t-best populations

Let $\Pi_1,\ldots,\Pi_k$ be k independent populations with Π_i associated with distribution function $F_i(x) = F(x-\theta_i)$, where θ_i is unknown for all $i = 1, 2, \ldots, k$. Denote by $\theta_{[1]} \le \theta_{[2]} \le \cdots \le \theta_{[k]}$ the true (unknown) ordering of the parameters. Let $t < k$, then we say that Π_i is among the t-best populations if $\theta_i \ge \theta_{[k-t+1]}$. We wish to select exactly t populations such that any of them is among the t-best populations. The problem will be formulated as follows:

Let $X = \{\underset{\sim}{x} = (x_1,\ldots,x_k) \mid -\infty < x_i < \infty$ for all $i = 1, 2, \ldots, k\}$ and $\Theta = \{\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_k) \mid -\infty < \theta_i < \infty$ for all $i = 1, 2, \ldots, k\}$. Also let $K = \{1,2,\ldots,k\}$, $T = \{1,2,\ldots,t\}$. Let $S = \{s \mid s : T \to K$ is 1 - 1 function and $s(i) < s(j)$ if $i < j\}$, then for each $s \in S$, $\{\Pi_{s(1)},\ldots,\Pi_{s(t)}\}$ will denote a possible choice of the set of t-test populations. It is clear that S contains $\binom{k}{t}$ elements, we will denote them by $s_1, s_2, \ldots, s_r$, where $r = \binom{k}{t}$. In view of the definitions given above, we use the Borel σ-field for Θ and X, and the discrete σ-field for S.

Definition 2.2.1. A measurable function $\phi : \mathcal{X} \times S \to [0,1]$ is a selection *rule if for each* $\underset{\sim}{x} \in \mathcal{X}$, *we have*

$$\sum_{s \in S} \phi(\underset{\sim}{x},s) = 1.$$

It is understood that $\phi(\underset{\sim}{x},s)$ is the conditional probability of selecting $\{\Pi_{s(1)},\ldots,\Pi_{s(k)}\}$, having observed $\underset{\sim}{x}$. For $1 \le i \le k$, let

$$S_{i1} = \{s \in S \mid i \in s(T)\}$$

$$S_{i2} = \{s \in S \mid i \notin s(T)\},$$

then $S_{i1}(S_{i2})$ is the collection of all subsets of size t which includes (does not include) Π_i. For each given ϕ, we have the following definition.

Definition 2.2.2. The k functions defined by

$$\delta_i(\underset{\sim}{x}) = \sum_{s \in S_{i1}} \phi(\underset{\sim}{x},s), \quad i = 1, 2, \ldots, k, \tag{2.2.1}$$

are the individual selection probabilities; $\delta_i(\underset{\sim}{x})$ is the conditional probability of including population Π_i in the selected subset having observed $\underset{\sim}{x}$. It follows that $\delta_i(\underset{\sim}{x})$ satisfies:

$$\text{(i)} \quad 0 \le \delta_i(\underset{\sim}{x}) \le 1 \text{ for all } 1 \le i \le k, \text{ and all } \underset{\sim}{x}. \tag{2.2.2}$$

$$\text{(ii)} \quad \sum_{i=1}^{k} \delta_i(\underset{\sim}{x}) = \sum_{i=1}^{k} \sum_{s \in S_{i1}} \phi(\underset{\sim}{x},s) = \sum_{s \in S} \sum_{i \in s(T)} \phi(\underset{\sim}{x},s)$$

$$= \sum_{s \in S} t\,\phi(\underset{\sim}{x},s) = t, \text{ for all } \underset{\sim}{x}. \tag{2.2.3}$$

Now, we have

Lemma 2.2.1. Given $\underset{\sim}{\delta}(\underset{\sim}{x}) = (\delta_1(\underset{\sim}{x}),\ldots,\delta_k(\underset{\sim}{x}))'$ satisfying (2.2.2), (2.2.3), there always exists at least one selection rule ϕ such that (2.2.1) holds.

Proof: (2.2.1) defines a simultaneous linear equation

$$A_{k\times r}\phi(\underset{\sim}{x})_{r\times 1} = \underset{\sim}{\delta}(\underset{\sim}{x})_{k\times 1} \tag{2.2.4}$$

where

$$\phi(\underset{\sim}{x}) = (\phi(\underset{\sim}{x},s_1),\ \phi(\underset{\sim}{x},s_2),\ldots,\ \phi(\underset{\sim}{x},s_r))'$$

$A = (a_{ij})$ is the matrix with

$$a_{ij} = \begin{cases} 0 & \text{if } i \notin s_j(T) \\ 1 & \text{if } i \in s_j(T) \end{cases} \qquad \text{for all } i = 1,2,\ldots,k.$$

It is understood that for $\underset{\sim}{x}$ fixed, $\underset{\sim}{\delta}(\underset{\sim}{x})$ is just a vector in $\mathbb{R}^k$ and $\phi(\underset{\sim}{x})$ is just a vector in $\mathbb{R}^r$. For simplicity, they will be denoted by $\underset{\sim}{v} = (v_1,v_2,\ldots,v_k)'$ and $\underset{\sim}{u} = (u_1,u_2,\ldots,u_r)'$, respectively. Now, consider $V = \{v \mid \sum_{i=1}^{k} v_i = t\} \cap [0,1]^k$, then V is a closed and bounded convex set. For $\underset{\sim}{v} \in V$, wlog, we can let $\underset{\sim}{v} = (1,1,\ldots,1,a_1,\ldots,a_\ell,0,\ldots,0)$ where $1 > a_1 \geq \ldots \geq a_\ell > 0$. If $\ell \neq 0$, let $\varepsilon_1 = 1 - a_1$, $\varepsilon_2 = a_\ell$ and $\varepsilon = \frac{1}{2}(\varepsilon_1 \wedge \varepsilon_2)$, then

$$\begin{aligned} \underset{\sim}{v} = {} & \tfrac{1}{2}(1,\ldots,1,a_1+\varepsilon,a_2-\varepsilon,\ldots,a_{2m-1}+\varepsilon,a_{2m}-\varepsilon,0,\ldots,0) + \\ & \tfrac{1}{2}(1,\ldots,1,a_1-\varepsilon,a_2+\varepsilon,\ldots,a_{2m-1}-\varepsilon,a_{2m}+\varepsilon,0,\ldots,0) \\ & \qquad\qquad \text{if } \ell = 2m, \end{aligned}$$

and

$$\underset{\sim}{v} = \frac{1}{2}(1,\ldots,1,a_1+\varepsilon,a_2-\varepsilon,\ldots,a_{2m+1}+\varepsilon,a_{2m+2}+\varepsilon,a_{2m+3}-2\varepsilon)$$

$$+ \frac{1}{2}(1,\ldots,1,a_1-\varepsilon,a_2+\varepsilon,\ldots,a_{2m+1}-\varepsilon,a_{2m+2}-\varepsilon,a_{2m+3}+2\varepsilon)$$

$$\text{if } \ell = 2m+3 .$$

Hence if $\ell \neq 0$, $\underset{\sim}{v}$ is not an extreme point of V. This shows the extreme points of V are the permutations of $(1,1,\ldots,1,0,\ldots,0)'$ (with t 1's), and they are just the columns of A. Since points in V can be expressed as a linear combination of its extreme points, this proves that equation (2.2.4) has at least a solution which is also a selection rule.

Lemma 2.2.1 does not exclude the possibility that more than one selection rules may have the same individual selection probabilities. For example, when $k = 4$, $t = 2$, let $\Pi_1 > \Pi_2 > \Pi_3 > \Pi_4$, and if

$$\phi_1 \text{ selects } \begin{cases} (\Pi_1,\Pi_2) & \text{with probability } \frac{1}{3} \\ (\Pi_2,\Pi_3) & \text{with probability } \frac{1}{3} \\ (\Pi_3,\Pi_4) & \text{with probability } \frac{1}{3} \end{cases}$$

$$\phi_2 \text{ selects } \begin{cases} (\Pi_1,\Pi_4) & \text{with probability } \frac{1}{3} \\ (\Pi_2,\Pi_3) & \text{with probability } \frac{2}{3} \end{cases} ,$$

then $\phi_1 \neq \phi_2$; but they have the same individual selection probabilities. ϕ_1 is better than ϕ_2 in the sense that ϕ_1 has $\frac{1}{3}$ chance to select the true 2-best populations. However, ϕ_2 is better than ϕ_1 in the sense that ϕ_2 always selects one of the 2-best populations. At this

stage, it is hard to judge which one of ϕ_1 and ϕ_2 is better than the other. For our convenience, we simply treat them as equivalent. Thus, we can consider decision rules in terms of the individual selection probabilities.

Let $D = \{\underset{\sim}{\delta} = (\delta_1,\ldots,\delta_k)' \mid 0 \le \delta_i(\underset{\sim}{x}) \le 1$ for $1 \le i \le k$, and $\sum_{i=1}^{k} \delta_i(\underset{\sim}{x}) = t\}$. For given $\varepsilon > 0$ and $s \in S$, let $\Theta_s = \{\underset{\sim}{\theta} \in \Theta \mid \min_{i \in s(T)} \theta_i \ge \max_{i \notin s(T)} \theta_i + \varepsilon\}$. Then, $\Theta = [\bigcup_{s \in S} \Theta_s] \cup \Theta_0$, where $\Theta_0 = \Theta \setminus (\bigcup_{s \in S} \Theta_s)$. Note that $\Theta_{s_i} \cap \Theta_{s_j} = \phi$ if $i \ne j$.

<u>Definition 2.2.3.</u> For any $s \in S$, Let $\lambda_s \in [0,1]$ be given and $\sum_{s \in S} \lambda_s \le 1$. Then

$$\Gamma = \{\tau \mid \tau \text{ is a prior distribution on } \Theta \text{ and } \int_{\Theta_s} d\tau(\theta) = \lambda_s \text{ for all } s \in S\}.$$

<u>Definition 2.2.4.</u> For any $\underset{\sim}{\theta} \in \Theta$ and $\underset{\sim}{\delta} \in D$, the loss function L is defined as

$$L(\underset{\sim}{\theta},\underset{\sim}{\delta}(\underset{\sim}{x})) = \sum_{s \in S} \sum_{j=1}^{k} L^{(s)}(\underset{\sim}{\theta},\delta_j(\underset{\sim}{x})),$$

where

$$L^{(s)}(\underset{\sim}{\theta},\delta_j(\underset{\sim}{x})) = \begin{cases} 0 & \text{for all } j \text{ if } \underset{\sim}{\theta} \notin \Theta_s \\ L_{s1}(1-\delta_j(\underset{\sim}{x})) & \text{for } j \in s(T),\ \underset{\sim}{\theta} \in \Theta_s \\ L_{s2}\,\delta_j(\underset{\sim}{x}) & \text{for } j \notin s(T),\ \underset{\sim}{\theta} \in \Theta_s \end{cases}$$

A similar loss function was considered by Gupta and Huang (1977). Let us further assume that Π_i has pdf $f_i(x) = f(x-\theta_i)$ and let $f_{\underset{\sim}{0}}(\underset{\sim}{x}) = \prod_{i=1}^{k} f_i(x_i)$. Then, we have

$$r(\tau,\underset{\sim}{\delta}) = \sum_{s\in S} \int_{\Theta_s} \int_{\mathbb{R}^k} \sum_{j=1}^{k} L^{(s)}(\underset{\sim}{\theta},\delta_j(\underset{\sim}{x}))f_{\underset{\sim}{\theta}}(\underset{\sim}{x})\, d\underset{\sim}{x}\, d\tau(\underset{\sim}{\theta})$$

$$= \sum_{s\in S} \int_{\Theta_s} \sum_{j\in s(T)} L_{s1} + L_{s2}\, E_{\underset{\sim}{\theta}}\Big[\sum_{j\notin s(T)} \delta_j(\underset{\sim}{X})\Big] - L_{s1}\, E_{\underset{\sim}{\theta}}\Big[\sum_{j\in s(T)} \delta_j(\underset{\sim}{X})\Big] d\tau(\underset{\sim}{\theta}). \tag{2.2.5}$$

Now, we can prove the following theorem:

Theorem 2.2.1. If for all $s \in S$, there exists a $\underset{\sim}{\theta}_s^* \in \Theta_s$ such that

$$\sup_{\underset{\sim}{\theta}\in\Theta_s} E_{\underset{\sim}{\theta}}\Big[\sum_{j\notin S(T)} \delta_j^0(\underset{\sim}{X})\Big] = E_{\underset{\sim}{\theta}_s^*}\Big[\sum_{j\notin S(T)} \delta_j^0(\underset{\sim}{X})\Big]$$

and

$$\inf_{\underset{\sim}{\theta}\in\Theta_s} E_{\underset{\sim}{\theta}}\Big[\sum_{j\in S(T)} \delta_j^0(\underset{\sim}{X})\Big] = E_{\underset{\sim}{\theta}_s^*}\Big[\sum_{j\in S(T)} \delta_j^0(\underset{\sim}{X})\Big] ,$$

where

$$\delta_j^0(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } N_j(\underset{\sim}{x}) < N_{[t]}(\underset{\sim}{x}) \\ r_j(\underset{\sim}{x}) & \text{if } N_j(\underset{\sim}{x}) = N_{[t]}(\underset{\sim}{x}), \quad \sum_{j=t_1+1}^{t_2} r_j(\underset{\sim}{x}) = t - t_1 \\ 0 & \text{if } N_j(\underset{\sim}{x}) > N_{[t]}(\underset{\sim}{x}) \end{cases} \tag{2.2.6}$$

and

$$N_{[1]}(\underset{\sim}{x}) < \dots < N_{[t_1+1]}(\underset{\sim}{x}) = \dots = N_{[t]}(\underset{\sim}{x}) = \dots = N_{[t_2]}(\underset{\sim}{x}) < \dots < N_{[k]}(\underset{\sim}{x})$$

is an ordered permutation of $N_j(\underset{\sim}{x})$, $j = 1, 2, \dots, k$, and

$$N_j(\underset{\sim}{x}) = M_{j2}(\underset{\sim}{x}) - M_{j1}(\underset{\sim}{x}) \text{ , where}$$

$$M_{j2}(\underset{\sim}{x}) = \sum_{s\in S_{j2}} L_{s2}\, \lambda_s\, f_{\underset{\sim}{\theta}_s^*}(\underset{\sim}{x}), \quad M_{j1}(\underset{\sim}{x}) = \sum_{s\in S_{j1}} L_{s1}\lambda_s\, f_{\underset{\sim}{\theta}_s^*}(\underset{\sim}{x}) \ .$$

Then, $\underset{\sim}{\delta}^0 = (\delta_1^0,\ldots,\delta_k^0)$ is a Γ-minimax rule.

Proof: Let τ_0 be such that $P_{\tau_0}[\underset{\sim}{\theta} = \underset{\sim}{\theta}_s^*] = \lambda_s$ for all $s \in S$, then for all $\underset{\sim}{\delta} \in D$,

$$\sup_{\tau\in\Gamma} r(\tau,\underset{\sim}{\delta}) \geq r(\tau_0,\underset{\sim}{\delta}) = \sum_{s\in S} \lambda_s \int_{\mathbb{R}^k} \sum_{j=1}^{k} L^{(s)}(\underset{\sim}{\theta}_s^*,\delta_j(\underset{\sim}{x}))\, f_{\underset{\sim}{\theta}_s^*}(\underset{\sim}{x})\, d\underset{\sim}{x}$$

$$= \sum_{j=1}^{k} \int_{\mathbb{R}^k} \Big(\sum_{s\in S_{j1}} + \sum_{s\in S_{j2}} \Big) [L^{(s)}(\underset{\sim}{\theta}_s^*,\delta_j(\underset{\sim}{x}))\, \lambda_s f_{\underset{\sim}{\theta}_s^*}(\underset{\sim}{x})]\, d\underset{\sim}{x}$$

$$= \sum_{j=1}^{k} \int_{\mathbb{R}^k} M_{j1}(\underset{\sim}{x}) + N_j(\underset{\sim}{x})\ \delta_j(\underset{\sim}{x})\ d\underset{\sim}{x}$$

$$\geq \sum_{j=1}^{k} \int_{\mathbb{R}^k} M_{j1}(\underset{\sim}{x}) + N_j(\underset{\sim}{x})\ \delta_j^0(\underset{\sim}{x})\ d\underset{\sim}{x}$$

$$= \sum_{s\in S} \lambda_s \int_{\mathbb{R}^k} \sum_{j=1}^{k} L^{(s)}(\underset{\sim}{\theta}_s^*,\delta_j^0(\underset{\sim}{x}))\, f_{\underset{\sim}{\theta}_s^*}(\underset{\sim}{x})\, d\underset{\sim}{x}$$

$$= \sum_{s\in S} \lambda_s \Big\{ \sum_{j\in s(T)} L_{s1} + L_{s2} E_{\underset{\sim}{\theta}_s^*}\Big[\sum_{j\notin s(T)} \delta_j^0(\underset{\sim}{X})\Big]$$

$$- L_{s1} E_{\underset{\sim}{\theta}_s^*}\Big[\sum_{j\in s(T)} \delta_j^0(\underset{\sim}{X})\Big] \Big\}$$

$$\geq \sum_{s\in S} \int_{\Theta_s} \sum_{j\in s(T)} L_{s1} + L_{s2} E_{\underset{\sim}{\theta}}\Big[\sum_{j\notin s(T)} \delta_j^0(\underset{\sim}{X})\Big]$$

$$- L_{s1} E_{\underset{\sim}{\theta}}\Big[\sum_{j\in s(T)} \delta_j^0(\underset{\sim}{X})\Big] d\tau(\underset{\sim}{\theta})$$

$$= r(\tau,\delta^0) \qquad \text{for all } \tau \in \Gamma \ .$$

So,

$$\sup_{\tau\in\Gamma} r(\tau,\underset{\sim}{\delta}) \geq \sup_{\tau\in\Gamma} r(\tau,\underset{\sim}{\delta}^0) .$$

This proves that $\underset{\sim}{\delta}^0$ is a Γ-minimax rule.

Let us take $\underset{\sim}{\theta}_s^* = (\theta_1,\theta_2,\ldots,\theta_k)$, where

$\theta_i = \theta_0+\varepsilon$ if $i \in s(T)$ and $\theta_i = \theta_0$ if $i \notin s(T)$. (2.2.7)

The constant θ_0 will be determined later. We would like to investigate $N_i(\underset{\sim}{x}) \leq N_j(\underset{\sim}{x})$ first. We find that

$$N_j(\underset{\sim}{x}) \geq N_i(\underset{\sim}{x})$$

if and only if

$$\sum_{s\in S_{i1}} (L_{s1}+L_{s2})\,\lambda_s f_{\underset{\sim}{\theta}_s^*}(\underset{\sim}{x}) \geq \sum_{s\in S_{j1}} (L_{s1}+L_{s2})\,\lambda_s f_{\underset{\sim}{\theta}_s^*}(\underset{\sim}{x}) .$$

If $(L_{s1}+L_{s2})\,\lambda_s = c$, where c is some constant, for all $s \in S$, then

$$N_j(\underset{\sim}{x}) \geq N_i(\underset{\sim}{x}) \iff \sum_{s\in S_{i1}\setminus S_{j1}} f_{\underset{\sim}{\theta}_s^*}(\underset{\sim}{x}) \geq \sum_{s\in S_{j1}\setminus S_{i1}} f_{\underset{\sim}{\theta}_s^*}(\underset{\sim}{x})$$

$$\iff \frac{f_{\theta_0+\varepsilon}(x_i)}{f_{\theta_0}(x_i)} \sum_{s\in S_{i1}\setminus S_{j1}} \prod_{\ell\in s(T)\setminus\{i\}} \frac{f_{\theta_0+\varepsilon}(x_\ell)}{f_{\theta_0}(x_\ell)}$$

$$\geq \frac{f_{\theta_0+\varepsilon}(x_j)}{f_{\theta_0}(x_j)} \sum_{s\in S_{j1}\setminus S_{i1}} \prod_{\ell\in s(T)\setminus\{j\}} \frac{f_{\theta_0+\varepsilon}(x_\ell)}{f_{\theta_0}(x_\ell)} .$$

But for all $s \in S_{i1}\setminus S_{j1}$, $(s(T)\setminus\{i\}) \cup \{j\} = s'(T)$ for some $s' \in S_{j1}\setminus S_{i1}$; and vise versa. So we get

$$\sum_{s\in S_{i1}\setminus S_{j1}} \prod_{\ell\in s(T)\setminus\{i\}} \frac{f_{\theta_0+\varepsilon}(x_\ell)}{f_{\theta_0}(x_\ell)} = \sum_{s\in S_{j1}\setminus S_{i1}} \prod_{\ell\in s(T)\setminus\{j\}} \frac{f_{\theta_0+\varepsilon}(x_\ell)}{f_{\theta_0}(x_\ell)} .$$

Hence,

$$N_i(\underset{\sim}{x}) \le N_j(\underset{\sim}{x}) \quad \text{iff} \quad g_{\theta_0}(x_i) \ge g_{\theta_0}(x_j) \tag{2.2.8}$$

where

$$g_{\theta_0}(x) = \frac{f_{\theta_0+\varepsilon}(x)}{f_{\theta_0}(x)} \quad .$$

Now, g_{θ_0} is increasing in x for any θ_0 if $f_\theta(x)$ has monotone likelihood ratio in x. Then,

$$x_i \ge x_j \implies N_i(x) \le N_j(\underset{\sim}{x}) \quad .$$

It is well known that if $X_1,\ldots,X_k$ are independent and the density $f_{\theta_i}(x)$ of X_i has MLR in x for all $i = 1,\ldots,k$, then $E_{\underset{\sim}{\theta}}[\delta(\underset{\sim}{X})]$ is increasing(decreasing) in θ_i if $\delta(\underset{\sim}{x})$ is increasing(decreasing) in x_i. (Lehmann(1959))

We can now state the main theorem of this section as follows:

Theorem 2.2.2. Let $X_1,\ldots,X_k$ be independent random variables. Assume that X_i has pdf $f_{\theta_i}(x) = f(x-\theta_i)$, which has MLR in x. Furthermore, if for all $s \in S$, $(L_{s1}+L_{s2})\lambda_s = c$, for some constant c, then $\underset{\sim}{\delta}^\star = (\delta_1^\star,\delta_2^\star,\ldots,\delta_k^\star)$ is a Γ-minimax rule, where

$$\delta_j^\star(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } x_j > x^{[t]} \\ \dfrac{t-t_1'}{t_2'-t_1'} & \text{if } x_j = x^{[t]} \\ 0 & \text{if } x_j < x^{[t]} \end{cases} \tag{2.2.9}$$

and

$$x^{[1]} > \ldots > x^{[t_1'+1]} = \ldots = x^{[t]} = \ldots = x^{[t_2']} > \ldots > x^{[k]}$$ is an

ordered permutation of $\underset{\sim}{x}$.

Proof: Let θ_s^* be defined by (2.2.7), then (2.2.8) holds. Now, we let $x_{i_j} = x^{[j]}$, then we have

$$g_{\theta_0}(x_{i_1}) \geq \ldots \geq g_{\theta_0}(x_{i_{t_1'+1}}) = \ldots = g_{\theta_0}(x_{i_t}) = \ldots = g_{\theta_0}(x_{i_{t_2'}}) \geq \ldots \geq g_{\theta_0}(x_{i_k}).$$

Hence, $N_{i_1}(\underset{\sim}{x}) \leq \ldots \leq N_{i_{t_1'+1}}(\underset{\sim}{x}) = \ldots = N_{i_t}(\underset{\sim}{x}) = \ldots = N_{i_{t_2'}}(\underset{\sim}{x}) \leq \ldots \leq N_{i_k}(\underset{\sim}{x})$.

Now, suppose that

$$N_{[1]}(\underset{\sim}{x}) < \ldots < N_{[t_1+1]}(\underset{\sim}{x}) = \ldots = N_{[t]}(\underset{\sim}{x}) \ldots = N_{[t_2]}(\underset{\sim}{x}) < \ldots < N_{[k]}(\underset{\sim}{x}) ,$$

then

$$t_1 \leq t_1' \leq t_2' \leq t_2 .$$

If in (2.2.6) we let

$$r_j(\underset{\sim}{x}) = \begin{cases} 1 & \text{for } j \text{ such that } x_{i_{t_1+1}} \geq x_j \geq x_{i_{t_1'}} \\[2ex] \dfrac{t-t_1'}{t_2'-t_1'} & \text{for } j \text{ such that } x_{i_{t_1'+1}} \geq x_j \geq x_{i_{t_2'}} \\[2ex] 0 & \text{for } j \text{ such that } x_{i_{t_2'+1}} \geq x_j \geq x_{i_{t_2}} \end{cases} ,$$

then $\delta_j^0(\underset{\sim}{x})$ reduces to $\delta_j^*(\underset{\sim}{x})$ as shown in (2.2.9).

Now, for any $s \in S$, let $i \notin s(T)$, consider

$$\underset{\sim}{x} = (x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_k)$$

and

$$\underset{\sim}{x}' = (x_1,\ldots,x_{i-1},x_i',x_{i+1},\ldots,x_k) \quad ,$$

where $x_i \le x_i'$, then for all $j \in s(T)$,

$$\delta_j^*(\underset{\sim}{x}) \geq \delta_j^*(\underset{\sim}{x}') \Longrightarrow \sum_{j\in s(T)} \delta_j^*(\underset{\sim}{x}) \geq \sum_{j\in s(T)} \delta_j^*(\underset{\sim}{x}') \quad ,$$

which shows that $\sum_{j\in s(T)} \delta_j^*(\underset{\sim}{x})$ is decreasing in x_i if $i \notin s(T)$. Hence $\sum_{j\notin s(T)} \delta_j^*(\underset{\sim}{x})$ is increasing in x_i if $i \notin s(T)$, because $\sum_{j\in s(T)} \delta_j^*(\underset{\sim}{x}) + \sum_{j\notin s(T)} \delta_j^*(\underset{\sim}{x}) = t$. Similarly, we can prove that $\sum_{j\notin s(T)} \delta_j^*(\underset{\sim}{x})$ is decreasing in x_i for $i \in s(T)$ and $\sum_{j\in s(T)} \delta_j^*(\underset{\sim}{x})$ is increasing in x_i for $i \in s(T)$. It follows that $E_{\underset{\sim}{\theta}}[\sum_{j\notin s(T)} \delta_j^*(\underset{\sim}{X})]$ is an increasing function of θ_i for $i \notin s(T)$ and is a decreasing function of θ_i for $i \in s(T)$. Hence

$$\sup_{\underset{\sim}{\theta}\in\Theta_s} E_{\underset{\sim}{\theta}}[\sum_{j\notin s(T)} \delta_j^*(\underset{\sim}{X})] = \sup_{-\infty<\theta_0<\infty} E_{\underset{\sim}{\theta}_s^*}[\sum_{j\notin s(T)} \delta_j^*(\underset{\sim}{X})] \quad .$$

Now,

$$x_j \gtreqless x^{[t]} \quad \text{iff} \quad y_j \gtreqless y^{[t]} \quad \text{where} \quad y_i = x_i - \theta_0$$

and the distribution of $X_j - \theta_0$ does not depend on θ_0 any more. This implies that $E_{\underset{\sim}{\theta}_s^*}[\sum_{j\notin s(T)} \delta_j^*(\underset{\sim}{X})]$ is independent of the choice of θ_0, so

$$\sup_{\underset{\sim}{\theta}\in\Theta_s} E_{\underset{\sim}{\theta}}[\sum_{j\notin s(T)} \delta_j^*(\underset{\sim}{X})] = E_{\underset{\sim}{\theta}_s^*}[\sum_{j\notin s(T)} \delta_j^*(\underset{\sim}{X})] \quad .$$

By the same argument,

$$\inf_{\underset{\sim}{\theta}\in\Theta_s} E_{\underset{\sim}{\theta}}[\sum_{j\in s(T)} \delta_j^*(\underset{\sim}{X})] = E_{\underset{\sim}{\theta}_s^*}[\sum_{j\in s(T)} \delta_j^*(\underset{\sim}{X})] \quad .$$

Hence, δ^* is a Γ-minimax rule by Theorem 2.2.1.

Remarks:

1. We are considering location parameters for continuous distribution, hence the probability of ties among x's is 0. This means that the natural selection rule (select the populations associated with the largest t ordered statistics among X's) is a Γ-minimax selection rule.

2. Assume $X_{i1},\dots,X_{in}$ are the observations from Π_i, and $\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$ is a sufficient statistic for θ_i. In this case, θ_i is still a location parameter for $\bar{X}_i$ and hence the Γ-minimax rule will select the populations associated with the t largest sample means. One such example is when $\Pi_i \sim N(\theta_i,\sigma^2)$, where σ^2 is known.

3. The condition $(L_{s1}+L_{s2})\lambda_s = c$ for all $s \in S$ holds if we let $L_{s1} = L_1$, $L_{s2} = L_2$ and $\lambda_s = \lambda$ for all $s \in S$, then Γ reduces to $\Gamma_\lambda = \{\tau| \int_{\Theta_s} d\tau(\underset{\sim}{\theta}) = \lambda$ for all $s \in S\}$. Γ_λ is a small class of prior distributions. But it is interesting to note that the Γ-minimax rule $\underset{\sim}{\delta}^*$ is actually independent of λ. So if we let $\Gamma_I = \bigcup_{\lambda\in I} \Gamma_\lambda$, where I is a arbitrary subset of the interval $J = [0,\frac{1}{\binom{k}{t}}]$, then $\underset{\sim}{\delta}^*$ is a Γ-minimax rule for $\Gamma = \Gamma_I$.

4. The loss function we used in this section (see Definition 2.2.4) satisfies the monotonicity and invariance properties of Eaton's paper (1967), and $f_{\underset{\sim}{\theta}}(\underset{\sim}{x})$ has the M-property (which is equivalent to MLR if X's are independent) so from Eaton's Theorem 4.1, $\underset{\sim}{\delta}^*$ is a Bayes rule wrt τ for any $\tau \in \Gamma' = \{\tau|\ \tau$ is an exchangeable prior distribution on $\Theta\}$. Then, $\underset{\sim}{\delta}^*$ is also a Γ-minimax rule for any $\Gamma \subseteq \Gamma'$.

5. It is easily seen that $\Gamma\subset\Gamma_J$, but $\Gamma\neq\Gamma_J$. To see this, let k=2, t=1, and let $\tau\sim N(\binom{0}{0},\begin{pmatrix}1&0\\0&2\end{pmatrix})$. Then, $\tau\notin\Gamma'$. However, $P_\tau[\theta_1-\theta_2\geq\varepsilon]$ $= P_\tau[\theta_2-\theta_1\geq\varepsilon] = P[X\geq\varepsilon]$, where $X\sim N(0,3)$, so $\tau\in\Gamma_J$. In this sense, our result is slightly stronger than Eaton's (1967).

6. If Π_i has a scale parameter θ_i, then $X_i\sim\frac{1}{\theta_i}f(\frac{x}{\theta_i})\, I_{(0,\infty)}(x)$ with $\theta_i > 0$. In this case, we might like to define

$$\Theta_s = \{\theta \mid (1-\varepsilon)\min_{i\in s(T)} \theta_i \geq \max_{i\notin s(T)} \theta_i\} \quad\text{and}\quad \Gamma = \{\tau \mid \int_{\Theta_s} d\tau(\theta) = \lambda \quad\text{for all}$$

$s\in S$, for some $\lambda\in J\}$. If we use the transformation $Y_i = \ell n X_i$, we get $Y_i\sim g(y-\eta_i)$, where $g(y) = e^y f(e^y)$ and $\eta_i = \ell n\theta_i$, also $\Gamma = \{\eta \mid \min_{i\in s(T)} \eta_i \geq (1+\varepsilon')\max_{i\notin s(T)} \eta_i\}$, where $1+\varepsilon' = \ell n\frac{1}{1-\varepsilon}$. Now, η_i is a location parameter, therefore, to choose the t-largest η_i's (hence the t-largest θ_i's), we will select those populations associated with the t-largest Y_i's (hence the t-largest X_i's).

In Section 2.4, we will see how Theorem 2.2.2 can be generalised if $X_1,X_2,\ldots,X_k$ are not assumed to be independent.

2.3 Complete ranking and simultaneous selection problems

Let $\Pi_1,\Pi_2,\ldots,\Pi_k$ be the same populations as described in Section 2.2 and $\theta_{[1]}\leq\ldots\leq\theta_{[k]}$ be the ordering or parameters. Let $R:\{\Pi_1,\Pi_2,\ldots,\Pi_k\}\to\{0,1,\ldots,k-1\}$ be a 1-1 function such as that $R(\Pi_i) = j-1$ iff $\theta_i = \theta_{[j]}$. $R(\Pi_i)$ is called the rank of Π_i. When $\theta_1,\theta_2,\ldots,\theta_k$ are unknown, the ranking problem is to identify each population with its rank. A simultaneous selection problem is to decide the t-best populations for all $1\leq t\leq k-1$ at the same time.

Definition 2.3.1. For ranking problem, let $A=\{\underset{\sim}{a} \mid \underset{\sim}{a}=(a(1),\ldots,a(k))'$ is a permutation of $(0,1,\ldots,k-1)'\}$. So when we take action $\underset{\sim}{a}\in A$, we mean population Π_i has rank $a(i)$. Notice there are $r=k!$ actions in A, which we denote by $\underset{\sim}{a}_1,\underset{\sim}{a}_2,\ldots,\underset{\sim}{a}_r$.

Definition 2.3.2. A measurable function $\delta: X \to A$ is called a ranking rule. A behaviorial ranking rule is a measurable function $\hat{\delta} : X \times A \to [0,1]$ such that $\hat{\delta}(\underset{\sim}{x},\cdot)$ is a probability measure on A. Then,
$\alpha_i(\underset{\sim}{x}) = \sum_{\ell=1}^{r} \hat{\delta}(x,\underset{\sim}{a}_\ell)\underset{\sim}{a}_\ell(i)$ is called the rank of Π_i generated by $\hat{\delta}$.

In the following, we would like to show the relation between the ranking problem and the simultaneous selection problem. A change of notation is necessary here. From now on, all the notations used in Section 2.2 will be added a sub-index t to specify that the selection is for the "t" best populations. For example,
$D_t = \{\underset{\sim}{\delta}_t = (\delta_{1t},\ldots,\delta_{kt})' \mid 0 \le \delta_{it}(x) \le 1, \sum_{i=1}^{k} \delta_{it}(\underset{\sim}{x}) = t\}$, for $1 \le t \le k-1$.

Definition 2.3.3. A general selection rule is a matrix $\underset{\approx}{\delta} = (\underset{\sim}{\delta}_1,\underset{\sim}{\delta}_2,\ldots,\underset{\sim}{\delta}_{k-1})$, where $\underset{\sim}{\delta}_j = (\delta_{1j},\ldots,\delta_{kj})' \in D_j$ for all $1 \le j \le k-1$. For any $\underset{\sim}{x} \in X$, $\underset{\approx}{\delta}(\underset{\sim}{x}) = [\delta_{ij}(\underset{\sim}{x})]_{k\times(k-1)}$, where $\delta_{ij}(\underset{\sim}{x})$ is the conditional probability of selecting Π_i as one of the j-best populations having observed $\underset{\sim}{x}$.

Definition 2.3.4. Let $\underset{\approx}{\delta} = [\delta_{ij}]$ be a general selection rule, then $\psi_i = \sum_{j=1}^{k-1} \delta_{ij}$ is called the rank of Π_i $(i=1,2,\ldots,k)$ generated by $\underset{\approx}{\delta}$.

Now, we can prove the following lemma to establish a relation between $\hat{\delta}(\underset{\sim}{x})$ and $\underset{\approx}{\delta}(x)$.

Lemma 2.3.1. Let $\hat{\delta}$ be a behaviorial ranking rule and $\alpha_i(\underset{\sim}{x}) = \sum_{\ell=1}^{r} \hat{\delta}(\underset{\sim}{x},\underset{\sim}{a}_\ell)\underset{\sim}{a}_\ell(i)$ be the rank of Π_i generated by $\hat{\delta}$, then there exists a general selection rule $\hat{\delta}$ such that the rank $\psi_i(\underset{\sim}{x}) = \sum_{j=1}^{k-1} \delta_{ij}(\underset{\sim}{x})$

is the same as $\alpha_i(\underset{\sim}{x})$, for all $1 \le i \le k$.

Proof: Since $\underset{\sim}{x}$ is fixed, we will use α_i for $\alpha_i(\underset{\sim}{x})$ to simplify notation. The same goes for ψ_i, $\hat{\delta}$, $\hat{\delta}_{ij}$ and $\underset{\approx}{\delta}$. Let $\hat{\delta}(\underset{\sim}{a}_\ell) = \beta_\ell$, then $0 \le \beta_\ell \le 1$ and $\sum_{\ell=1}^{r} \beta_\ell = 1$. Wlog, let $a_1 \in A$ be such that $a_1(i) = i - 1$ for all $1 \le i \le k$. Then, for all $a_\ell \in A$, $a_\ell = P_\ell a_1$, where P_ℓ is some permutation matrix $(2 \le \ell \le r)$. Now consider

$$\underset{\approx}{\delta}_1 = \begin{pmatrix} 0 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 0 & \cdot & & \cdot \\ 0 & 1 & & & \cdot \\ 1 & 1 & \cdot & \cdot & 1 \end{pmatrix}_{k\times(k-1)}, \quad \underset{\sim}{1}_{k-1} = \begin{pmatrix} 1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix}_{k-1},$$

then it is easy to check $\underset{\approx}{\delta}_1 1_{k-1} = \underset{\sim}{a}_1$, hence $P_\ell \underset{\approx}{\delta}_1 \underset{\sim}{1}_{k-1} = \underset{\sim}{a}_\ell$. Now, we define $\underset{\approx}{\delta} = \sum_{\ell=1}^{r} \beta_\ell P_\ell \underset{\approx}{\delta}_1$, where $P_1 = I_{k\times k}$. If we let $P_\ell = [p_{ij}^{(\ell)}]_{k\times k}$ and $\underset{\approx}{\delta} = [\delta_{ij}]_{k\times(k-1)}$, then we have

$$0 \le \delta_{ij} = \sum_{\ell=1}^{r} \beta_\ell \sum_{m=k-j+1}^{k} p_{im}^{(\ell)} \le \sum_{\ell=1}^{r} \beta_\ell = 1$$

and

$$\underset{\sim}{1}_k' \underset{\approx}{\delta} = \sum_{\ell=1}^{r} \beta_\ell \underset{\sim}{1}_k' P_\ell \underset{\approx}{\delta}_1 = \sum_{\ell=1}^{r} \beta_\ell \underset{\sim}{1}_k' \underset{\approx}{\delta}_1 = \underset{\sim}{1}_k' \underset{\approx}{\delta}_1 = (1,2,\ldots,k-1).$$

This proves $\sum_{i=1}^{k} \delta_{ij} = j$, hence $\underset{\sim}{\delta}_j \in D$, where $\underset{\sim}{\delta}_j = (\delta_{1j},\ldots,\delta_{kj})'$. It follows that $\underset{\approx}{\delta}$ is a general selection rule by Definition 2.3.3. Finally,

$$\underset{\approx}{\delta} \underset{\sim}{1}_{k-1} = (\psi_1,\ldots,\psi_k)' = \sum_{\ell=1}^{r} \beta_\ell P_\ell \underset{\approx}{\delta}_1 \underset{\sim}{1}_{k-1} = \sum_{\ell=1}^{r} \beta_\ell \underset{\sim}{a}_\ell$$

$$= \sum_{\ell=1}^{r} \hat{\delta}(\underset{\sim}{a}_\ell) \underset{\sim}{a}_\ell = (\alpha_1,\ldots,\alpha_k)',$$

hence $\psi_i = \alpha_i$ for all $1 \le i \le k$.

We can also prove that given a general selection rule $\underset{\approx}{\delta}$, there exists a behaviorial ranking rule $\hat{\delta}$ such that $\sum_{\ell=1}^{r} \hat{\delta}(\underset{\sim}{a}_\ell)\underset{\sim}{a}_\ell = \underset{\approx}{\delta}\underset{\sim}{1}_{k-1}$. The proof is very similar to the proof of Lemma 2.2.1. We consider that $V = \{\underset{\approx}{\delta}\underset{\sim}{1}_{k-1} \mid \underset{\approx}{\delta}$ is a general selection rule$\}$, then V is a closed, bounded convex set in $\mathbb{R}^k$. Now, if $\underset{\approx}{\delta} = [\delta_{ij}]$, but δ_{ij} are not all 0's and 1's, then $\underset{\approx}{\delta}\underset{\sim}{1}_{k-1}$ is not an extreme point. It turns out that the extreme points of V are $P_\ell \underset{\approx}{\delta}\underset{\sim}{1}_{k-1} = \underset{\sim}{a}_\ell$, so for all $\underset{\approx}{\delta}\underset{\sim}{1}_{k-1} \in V$, $\underset{\approx}{\delta}\underset{\sim}{1}_{k-1} = \sum_{\ell=1}^{r} \beta_\ell \underset{\sim}{a}_\ell$. Now, set $\hat{\delta}(\underset{\sim}{a}_\ell) = \beta_\ell$, thus completes the proof.

From the above discussion, the rank of Π_i generated by behaviorial ranking rule $\hat{\delta}$ or by general selection rule $\underset{\approx}{\delta}$ can be treated as equivalent. In the follwoing, we will consider the ranking problem through the general selection rule $\underset{\approx}{\delta}$, i.e., we would like to select the t-best populations for $1 \le t \le k - 1$ simultaneously and hence to rank the populations in some order.

Let $D = \{\underset{\approx}{\delta} \mid \underset{\approx}{\delta}$ is a general selection rule$\}$. The notation $L_t(\underset{\sim}{\theta},\underset{\sim}{\delta}_t(\underset{\sim}{x}))$ and $r_t(\tau,\underset{\sim}{\delta}_t)$ wil mean the same thing as $L(\underset{\sim}{\theta},\underset{\sim}{\delta}(\underset{\sim}{x}))$ and $r(\tau,\underset{\sim}{\delta})$ in Section 2.2. Because t is a variable rather than a fixed integer in this section, a sub-index t is added to make the notations clear.

An intuitive loss function for the simultaneous selection problem is $\ell(\underset{\sim}{\theta},\underset{\approx}{\delta}(x)) = \sum_{t=1}^{k-1} L_t(\underset{\sim}{\theta},\underset{\sim}{\delta}_t(\underset{\sim}{x}))$. Since the loss is additive, so by Lemma 1.7.1 and Theorem 4.1 of Eaton (1967), $\underset{\approx}{\delta}^* = (\underset{\sim}{\delta}_1^*,\ldots,\underset{\sim}{\delta}_{k-1}^*)$ is Bayes rule for any exchangeable prior distribution, where $\underset{\sim}{\delta}_t^*$ is as given in (2.2.9) for $1 \le t \le k - 1$. Then

$$\sup_{\tau\in\Gamma'} r(\tau,\underset{\approx}{\delta}^*) \le \sup_{\tau\in\Gamma'} r(\tau,\underset{\sim}{\delta}) \quad \text{for all} \quad \underset{\approx}{\delta} \in D,$$

i.e., $\underset{\approx}{\delta}^*$ is a Γ-minimax rule for $\Gamma = \Gamma'$. However, when we use $\ell(\underset{\sim}{\theta},\underset{\approx}{\delta}(\underset{\sim}{x}))$ as our loss function, we find that there is no indifference zone for $\underset{\approx}{\delta}$; but for each t $(1 \le t \le k - 1)$, $\underset{\approx}{\delta}_t$ has its own indifference zone. If we feel we should not be penalized in the problem of selecting the t-best populations (t fixed) when the t-best populations are not distinguishable from the others, then neither should we be penalized in the simultaneous selection problem if any two populations are not distinguishable from each other. Thus, we need to have an indifference zone for $\underset{\approx}{\delta}$, which is done as follows: Let

$$C = \{c \mid c \text{ is a permutation on } \{1,2,\ldots,k\}\},$$

and for $c \in C$, let

$$\Theta_c = \{\underset{\sim}{\theta} \mid \theta_{c(1)} > \theta_{c(2)} > \cdots > \theta_{c(k)} \text{ and } \theta_{c(i)} - \theta_{c(i+1)} \ge \varepsilon \text{ for } 1 \le i \le k-1\},$$

then

$$\Theta_0 = \Theta \setminus \bigcup_{c \in C} \Theta_c = \{\underset{\sim}{\theta} \mid \min_{1 \le i \le j \le k} |\theta_i - \theta_j| < \varepsilon\}$$ serves as

an indifference zone. Now let $C_{ij} = \{c \mid i = c(j)\}$, then for all $c \in C_{ij}$ and $\underset{\sim}{\theta} \in \Theta_c$, $\theta_i = \theta_{c(j)}$ is true, i.e., θ_i is the $j^{\underline{th}}$ largest parameter. Note that $C = \bigcup_{j=1}^{k} C_{ij}$, for all $1 \le i \le k$. For $1 \le t \le k$, $1 \le i \le k$, let $G_{it} = \bigcup_{j=1}^{t} C_{ij}$, $B_{it} = \bigcup_{j=t+1}^{k} C_{ij}$, then for $\underset{\sim}{\theta} \in \Theta_c$ and $c \in G_{it}$ (B_{it} respectively), we have θ_i is (is not) one of the t largest parameters. Now, we can define the loss function as:

<u>Definition 2.3.5.</u> For any $\underset{\sim}{\theta} \in \Theta$ and $\underset{\sim}{\delta} \in D$, let

$$L(\underset{\sim}{\theta},\underset{\approx}{\delta}) = \sum_{t=1}^{k-1} \ell_t(\underset{\sim}{\theta},\underset{\sim}{\delta}_t(\underset{\sim}{x})),$$

where

$$\ell_t(\underset{\sim}{\theta},\underset{\sim}{\delta}_t(\underset{\sim}{x})) = \sum_{i=1}^{k} \sum_{c \in C} L_t^{(c)}(\underset{\sim}{\theta},\delta_{it}(\underset{\sim}{x}))$$

and

$$L_t^{(c)}(\underset{\sim}{\theta},\delta_{it}(\underset{\sim}{x})) = \begin{cases} 0 & \text{if } \underset{\sim}{\theta} \notin \Theta_c \\ L_1(1-\delta_{it}(\underset{\sim}{x})) & \text{if } c \in G_{it} \text{ and } \underset{\sim}{\theta} \in \Theta_c \\ L_2\delta_{it}(\underset{\sim}{x}) & \text{if } c \in B_{it} \text{ and } \underset{\sim}{\theta} \in \Theta_c \end{cases}.$$

Let $\Gamma_\lambda = \{\tau | \int_{\Theta_c} d\tau(\underset{\sim}{\theta}) = \lambda$ for all $c \in C\}$. We see that Θ_0 is an indifference zone for $\underset{\approx}{\delta}(\underset{\sim}{x})$. Now, we can prove Theorem 2.3.1 which will be used to find a simultaneous Γ-minimax selection rule.

<u>Theorem 2.3.1.</u> If for all $c \in C$, there exists a $\underset{\sim}{\theta}_c^* \in \Theta_c$ such that

$$\sup_{\underset{\sim}{\theta} \in \Theta_c} E_{\underset{\sim}{\theta}}\Big[\sum_{j=t+1}^{k} \delta^0_{c(j)t}(\underset{\sim}{X})\Big] = E_{\underset{\sim}{\theta}_c^*}\Big[\sum_{j=t+1}^{k} \delta^0_{c(j)t}(\underset{\sim}{X})\Big] \tag{2.3.1}$$

and

$$\inf_{\underset{\sim}{\theta} \in \Theta_c} E_{\underset{\sim}{\theta}}\Big[\sum_{j=1}^{t} \delta^0_{c(j)t}(\underset{\sim}{X})\Big] = E_{\underset{\sim}{\theta}_c^*}\Big[\sum_{j=1}^{t} \delta^0_{c(j)t}(\underset{\sim}{X})\Big],$$

where

$$\delta^0_{it}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } N_{it}(\underset{\sim}{x}) < N_{[t]t}(x) \\ r_{it}(\underset{\sim}{x}) & \text{if } N_{it}(\underset{\sim}{x}) = N_{[t]t}(\underset{\sim}{x}), \text{ with } \sum_{i=t_1+1}^{t_2} r_{it}(\underset{\sim}{x}) = t-t_1 \\ 0 & \text{if } N_{it}(\underset{\sim}{x}) > N_{[t]t}(\underset{\sim}{x}) \end{cases} \tag{2.3.2}$$

and

$$N_{[1]t}(\underset{\sim}{x}) < \ldots < N_{[t_1]t}(\underset{\sim}{x}) = \ldots = N_{[t]t}(\underset{\sim}{x}) = \ldots = N_{[t_2]t}(\underset{\sim}{x}) < \ldots < N_{[k]t}(\underset{\sim}{x})$$

where

$$N_{it}(\underset{\sim}{x}) = L_2 \sum_{c \in B_{it}} f_{\underset{\sim}{\theta}_c^*}(\underset{\sim}{x}) - L_1 \sum_{c \in G_{it}} f_{\underset{\sim}{\theta}_c^*}(\underset{\sim}{x}) \ .$$

If τ_0 is defined as $P_{\tau_0}[\underset{\sim}{\theta} = \underset{\sim}{\theta}_c^*] = \lambda$ for all $c \in C$, then we have $r_t(\tau_0, \underset{\sim}{\delta}_t) \geq r_t(\tau_0, \underset{\sim}{\delta}_t^0) \geq r_t(\tau, \underset{\sim}{\delta}_t^0)$ for all $\underset{\sim}{\delta}_t \in D_t$ and $\tau \in \Gamma_\lambda$.

Proof: The proof is similar to that of Theorem 2.2.1, so we only write down the main steps and skip the details.

$$\begin{aligned}
r_t(\tau_0, \delta_t) &= K_t + \lambda \sum_{i=1}^{k} \int_{\mathbb{R}^k} N_{it}(\underset{\sim}{x})\, \delta_{it}(\underset{\sim}{x})\, d\underset{\sim}{x} \\
&\geq K_t + \lambda \sum_{i=1}^{k} \int_{\mathbb{R}^k} N_{it}(\underset{\sim}{x})\, \delta_{it}^0(\underset{\sim}{x})\, d\underset{\sim}{x} \\
&= r_t(\tau_0, \underset{\sim}{\delta}_t^0) \\
&= \lambda \sum_{c \in C} \{L_1 t + L_2 E_{\underset{\sim}{\theta}_c^*}[\sum_{j=t+1}^{k} \delta_{c(j)t}^0(\underset{\sim}{X})] \\
&\qquad - L_1 E_{\underset{\sim}{\theta}_c^*}[\sum_{j=1}^{t} \delta_{c(j)t}^0(\underset{\sim}{X})]\} \\
&\geq \sum_{c \in C} \int_{\Theta_c} L_1 t + L_2 E_{\underset{\sim}{\theta}}[\sum_{j=t+1}^{k} \delta_{c(j)t}^0(\underset{\sim}{X})] \\
&\qquad - L_1 E_{\underset{\sim}{\theta}}[\sum_{j=1}^{t} \delta_{c(j)t}^0(\underset{\sim}{X})]\, d\tau(\theta) \\
&= r_t(\tau, \underset{\sim}{\delta}_t^0) \quad \text{for all } \tau \in \Gamma_\lambda
\end{aligned}$$

where

$$K_t = \lambda \sum_{i=1}^{k} \int_X L_1 \sum_{c \in G_{it}} f_{\underset{\sim}{\theta}_c^*}(\underset{\sim}{x})\, d\underset{\sim}{x} \quad .$$

This completes the proof.

Now, for all $c \in C$, we define $\underset{\sim}{\theta}_c^* = (\theta_1, \theta_2, \ldots, \theta_k)' \in \Theta_c$ as

$$\theta_{c(j)} = \theta_0 - j\varepsilon \quad \text{for all} \quad 1 \le j \le k. \tag{2.3.3}$$

Again, θ_0 will be determined later. We would like to examine $N_{it}(\underset{\sim}{x}) \le N_{jt}(\underset{\sim}{x})$. We find

$$N_{it}(\underset{\sim}{x}) \le N_{jt}(\underset{\sim}{x}) \iff \sum_{c \in G_{it}} f_{\underset{\sim}{\theta}_c^*}(\underset{\sim}{x}) \ge \sum_{c \in G_{jt}} f_{\underset{\sim}{\theta}_c^*}(\underset{\sim}{x})$$

$$\iff \sum_{c \in G_{it} \setminus G_{jt}} f_{\underset{\sim}{\theta}_c^*}(\underset{\sim}{x}) \ge \sum_{c \in G_{jt} \setminus G_{it}} f_{\underset{\sim}{\theta}_c^*}(\underset{\sim}{x}) \quad .$$

Now, for all $c \in G_{it} \setminus G_{jt}$, $i = c(i')$ for some $1 \le i' \le t$ and $j = c(j')$ for some $t + 1 \le j' \le k$. If we let c' be such that $c'(\ell) = c(\ell)$ if $\ell \ne i'$, $\ell \ne j'$, and $c'(i') = j$, $c'(j') = i$, then $c' \in G_{jt} \setminus G_{it}$. The correspondence $c \leftrightarrow c'$ is 1 - 1 between $G_{it} \setminus G_{jt}$ and $G_{jt} \setminus G_{it}$. So if we let

$$g_c(\underset{\sim}{x}) = f_{\underset{\sim}{\theta}_c^*}(\underset{\sim}{x}) - f_{\underset{\sim}{\theta}_{c'}^*}(\underset{\sim}{x}) \quad \text{for all} \quad c \in G_{it} \setminus G_{jt},$$

we have

$$N_{it}(\underset{\sim}{x}) \le N_{jt}(\underset{\sim}{x}) \iff \sum_{c \in G_{it} \setminus G_{jt}} g_c(\underset{\sim}{x}) \ge 0.$$

Now,

$$g_c(\underset{\sim}{x}) = [\prod_{\ell\neq i',\ell\neq j'} f_{\theta_0-\ell\varepsilon}(x_\ell)]\,[f_{\theta_0-i'\varepsilon}(x_i)\, f_{\theta_0-j'\varepsilon}(x_j)$$

$$- f_{\theta_0-i'\varepsilon}(x_j)\, f_{\theta_0-j'\varepsilon}(x_i)] \;.$$

If $f_{\theta_i}(x)$ has MLR for all $1 \le i \le k$, then we have

$$x_i \ge x_j \implies g_c(\underset{\sim}{x}) \ge 0 \quad \text{for all} \quad c \in G_{it}\setminus G_{jt} \,,$$

$$\implies N_{it}(\underset{\sim}{x}) \le N_{jt}(\underset{\sim}{x}) \quad . \tag{2.3.4}$$

<u>Theorem 2.3.2.</u> Let $X_1, X_2, \ldots, X_k$ be independent random variables, where X_i has pdf $f_{\theta_i}(x) = f(x-\theta_i)$ which has MLR in x. Let

$$x^{[1]} > \ldots > x^{[t_1'+1]} = \ldots = x^{[t]} = \ldots = x^{[t_2']} > \ldots > x^{[k]}$$

be an ordered permutation of $\underset{\sim}{x}$. If

$$\delta^*_{it}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if} \quad x_i > x^{[t]} \\ \dfrac{t-t_1'}{t_2'-t_1'} & \text{if} \quad x_i = x^{[t]} \\ 0 & \text{if} \quad x_i < x^{[t]} \;, \end{cases} \tag{2.3.5}$$

then we have $\underset{\approx}{\delta}^*(\underset{\sim}{x}) = [\delta^*_{it}(\underset{\sim}{x})]_{k\times(k-1)}$ is a Γ-minimax simultaneous selection rule in D for $\Gamma = \Gamma_\lambda$.

<u>Proof</u>: Let $\underset{\sim}{\theta}_c^*$ be defined by (2.3.3), then by (2.3.4) we see that (2.3.5)can be considered as a special case of (2.3.2), as was shown in the proof of Theorem 2.2.2. Also, by an argument similar to that in the proof of Theorem 2.2.2, we can prove that (2.3.1) holds for any choice of θ_0. So by Theorem 2.3.1, we get

$$\inf_{\underset{\sim}{\delta}_t \in D_t} r_t(\tau_0, \underset{\sim}{\delta}_t) \geq r_t(\tau_0, \underset{\sim}{\delta}_t^*) \geq \sup_{\tau \in \Gamma_\lambda} r_t(\tau, \underset{\sim}{\delta}_t^*)$$

for all $1 \leq t \leq k-1$. Hence,

$$\inf_{\underset{\approx}{\delta} \in D} r(\tau_0, \underset{\approx}{\delta}) = \inf_{\underset{\approx}{\delta} \in D} \sum_{t=1}^{k} r_t(\tau_0, \underset{\sim}{\delta}_t) \geq \sum_{t=1}^{k} \inf_{\underset{\sim}{\delta}_t \in D_t} r_t(\tau_0, \underset{\sim}{\delta}_t^*)$$

$$\geq \sum_{t=1}^{k} r_t(\tau_0, \underset{\sim}{\delta}_t^*) = r(\tau_0, \underset{\approx}{\delta}^*) \quad .$$

Now,

$$\sup_{\tau \in \Gamma_\lambda} r(\tau, \underset{\approx}{\delta}^*) = \sup_{\tau \in \Gamma_\lambda} \sum_{i=1}^{k} r_t(\tau, \underset{\sim}{\delta}_t^*) \leq \sum_{i=1}^{k} \sup_{\tau \in \Gamma_\lambda} r_t(\tau, \underset{\sim}{\delta}_t^*)$$

$$\leq \sum_{i=1}^{k} r_t(\tau_0, \underset{\sim}{\delta}_t^*) = r(\tau_0, \underset{\approx}{\delta}^*) \leq \inf_{\underset{\approx}{\delta} \in D} r(\tau_0, \underset{\approx}{\delta})$$

$$\leq r(\tau_0, \underset{\approx}{\delta}) \leq \sup_{\tau \in \Gamma_\lambda} r(\tau, \underset{\approx}{\delta})$$

for all $\underset{\approx}{\delta} \in D$. So $\underset{\approx}{\delta}^*$ is a Γ-minimax simultaneous selection rule for $\Gamma = \Gamma_\lambda$.

<u>Corollary 2.3.1.</u> Let $I \subseteq [0, \frac{1}{k!}]$ and $\Gamma_I = \bigcup_{\tau \in I} \Gamma_\lambda$, then $\underset{\approx}{\delta}^*$ is a Γ-minimax rule for $\Gamma = \Gamma_I$.

Recall that our main purpose in doing the simultaneous selection is to rank the populations. When there are no ties among x_i's, i.e.

$$x_{i_1} < x_{i_2} < \ldots < x_{i_k} ,$$

then the rank $\psi_{i_j}(\underset{\sim}{x})$ generated by $\underset{\approx}{\delta}^*(\underset{\sim}{x})$ is $j - 1$, and hence Π_{i_j} has rank $j - 1$. If ties occur, wlog, let us assume that

$$x_1 > \ldots > x_{t_1+1} = \ldots = x_{t_1+d_1} > \ldots > x_{t_2+1} = \ldots = x_{t_2+d_2} > \ldots > x_{t_m+1} = \ldots$$

$$= x_{t_m+d_m} > \ldots > x_k ,$$

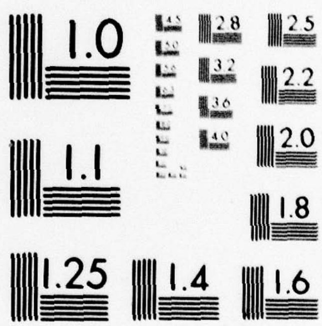
then

$$\underset{\approx}{\delta}^*(\underset{\sim}{x}) = \left[\begin{array}{ccccccccc}
1 & \cdots & 1 & & & & & & \\
 & \ddots & \vdots & & & & & \text{\Large 1} & \\
 & & 1 & & & & & & \\
 & & & \frac{1}{d_1} & \frac{2}{d_1} \;\cdots\; \frac{d_1}{d_1} & & & & \\
 & & & \vdots & \vdots \qquad\;\; \vdots & & & & \\
 & & & \frac{1}{d_1} & \frac{2}{d_1} \;\cdots\; \frac{d_1}{d_1} & & & & \\
 & & & & & 1 \cdots 1 & & & \\
 & & & & & \ddots \;\vdots & & & \\
 & & & & & 1 & & & \\
 & \text{\Large 0} & & & & & \frac{1}{d_2} \;\; \frac{2}{d_2} \;\cdots\; \frac{d_2}{d_2} & & \\
 & & & & & & \vdots \quad\; \vdots \qquad\;\; \vdots & & \\
 & & & & & & \frac{1}{d_2} \;\; \frac{2}{d_2} \;\cdots\; \frac{d_2}{d_2} & & \\
 & & & & & & & \ddots &
\end{array}\right]_{k\times(k-1)} .$$

AD-A073 574 PURDUE UNIV LAFAYETTE IND DEPT OF STATISTICS F/G 12/1
SOME CONTRIBUTIONS TO GAMMA-MINIMAX AND EMPIRICAL BAYES SELECTI--ETC(U)
AUG 79 P HSIAO N00014-75-C-0455
UNCLASSIFIED MMS-79-16 NL

2 OF 2
AD
A073574

END
DATE
FILMED
10-79
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

So,

$$\psi_i(\underset{\sim}{x}) = \sum_{t=1}^{k-1} \delta^*_{it}(\underset{\sim}{x})$$

$$= \begin{cases} k-i & \text{if } t_j + d_j < i < t_{j+1}+1 \text{ where } t_0=d_0=0 \\ \dfrac{1+d_j}{2} + k-1-t_j-d_j & \text{if } t_j + 1 \le i \le t_j + d_j. \end{cases}$$

$\psi_i(\underset{\sim}{x})$ is called midrank of $\underset{\sim}{x}$ and this justifies why the midranks for tied data should be used for rank test. For use of midranks for tied data, see Lehmann (1975).

2.4 Γ-minimax rules for hypothesis testing in a multivariate case.

We start with a result in Lehmann (1955), which we state as a lemma without a proof.

Definition 2.4.1. When $\underset{\sim}{x} = (x_1,\ldots,x_k)'$, $\underset{\sim}{x}'= (x_1',\ldots,x_k')'$, we define $\underset{\sim}{x} \le \underset{\sim}{x}'$ iff $x_i \le x_i'$ for all $1 \le i \le k$. A measurable set S is increasing iff $\underset{\sim}{x} \in S$ and $\underset{\sim}{x} \le \underset{\sim}{x}'$ implies $\underset{\sim}{x}' \in S$.

Definition 2.4.2. A family of distribution $\{F_{\underset{\sim}{\theta}}(\underset{\sim}{x})\}_{\underset{\sim}{\theta}\in\Theta}$ is said to have stochastically increasing property (SIP) iff when $\underset{\sim}{\theta} \le \underset{\sim}{\theta}'$ and S is an increasing set, we have $\int_S dF_{\underset{\sim}{\theta}}(\underset{\sim}{x}) \le \int_S dF_{\underset{\sim}{\theta}'}(\underset{\sim}{x})$.

An example of SIP family is when $F_{\underset{\sim}{\theta}}(\underset{\sim}{x}) = F(\underset{\sim}{x}-\underset{\sim}{\theta})$, i.e., $\underset{\sim}{\theta}$ is a location parameter. The following lemma is due to Lehmann (1955).

Lemma 2.4.1. Let $\{F_{\theta}(x)\}_{\theta \in \Theta}$ be a family of distribution with SIP. If δ is a real-valued function such that $\delta(x) \le \delta(x')$ for $x \le x'$, then $E_{\theta}[\delta(X)] \le E_{\theta'}[\delta(X)]$ for $\theta \le \theta'$.

When θ is a location parameter, the above lemma can be generalized to Lemma 2.4.2.

Lemma 2.4.2. Let $\{F(x-\theta)\}_{\theta \in \Theta}$ be a class of distribution. If δ is a real-valued function such that $\delta(x + ta) \ge \delta(x + sa)$ for $t \ge s$, then so is $E_{\theta + ta}[\delta(X)] \ge E_{\theta + sa}[\delta(X)]$, where a is an arbitrary vector in $\mathbb{R}^k$.

Proof: Let A be any non-singular matrix with a as its first column. Let $Y = A^{-1}X$ and $\hat{\delta}(x) = \delta(Ax)$, then when $X \sim f(x-\theta)$, we have $Y \sim c\hat{f}(x-\eta)$ where $c = |\det A|$, $\hat{f}(x) = f(Ax)$ and $\eta = A^{-1}\theta$. Also, we let $g(\theta) = E_{\theta}[\delta(X)]$ and $\hat{g}(\eta) = E_{\eta}[\hat{\delta}(Y)]$. Now,

$$\hat{\delta}(x + te_1) = \delta(Ax + tAe_1) = \delta(Ax + ta)$$

$$\ge \delta(Ax + sa) = \hat{\delta}(x + se_1)$$

for $t \ge s$, so $\hat{\delta}$ is increasing in its first component. Since η is the location parameter of Y, this implies $\hat{g}(\eta + te_1) \ge \hat{g}(\eta + se_1)$ if $t \ge s$. But

$$\begin{aligned} g(\theta) &= \int \delta(x)\, f(x - \theta)\, dx \\ &= \int \delta(Ax)\, f(Ax - AA^{-1}\theta)\, |\det A|\, dx \\ &= c \int \hat{\delta}(x)\, \hat{f}(x - A^{-1}\theta)\, dx \\ &= \hat{g}(A^{-1}\theta), \end{aligned}$$

hence,

$$g(\underset{\sim}{\theta} + t\underset{\sim}{a}) = \hat{g}(A^{-1}\underset{\sim}{\theta} + tA^{-1}\underset{\sim}{a}) = \hat{g}(A^{-1}\underset{\sim}{\theta} + t\underset{\sim}{e}_1)$$

$$\geq \hat{g}(A^{-1}\underset{\sim}{\theta} + s\underset{\sim}{e}_1) = g(\underset{\sim}{\theta} + s\underset{\sim}{a}) \quad \text{for all} \quad t \geq s.$$

This completes the proof.

Remarks:

1. One may notice that if $F_{\underset{\sim}{\theta}}(\underset{\sim}{x}) = F(\underset{\sim}{x}-\underset{\sim}{\theta})$, then Lemma 2.4.1 is an immediate result of Lemma 2.4.2.

2. If both $\delta(\underset{\sim}{x})$ and $g(\underset{\sim}{\theta})$ are differentiable, then we have
$$\sum_{i=1}^{k} a_i \frac{\partial}{\partial x_i} \delta(\underset{\sim}{x}) \geq 0 \Longrightarrow \sum_{i=1}^{k} a_i \frac{\partial}{\partial \theta_i} g(\underset{\sim}{\theta}) \geq 0 \quad \text{for any } \underset{\sim}{a}.$$

Example 2.4.1. Let the random variable X has pdf $f(\underset{\sim}{x}-\underset{\sim}{\theta}) = h(\underset{\sim}{x})\, c(\underset{\sim}{\theta})\, e^{\underset{\sim}{\theta}'\underset{\sim}{x}}$, where $\underset{\sim}{\theta} \in \mathbb{R}^k$ is unknown. Also let $\underset{\sim}{\beta}$ be any vector in $\mathbb{R}^k$. We want to test

$$H_0 : \underset{\sim}{\beta}'\underset{\sim}{\theta} \geq c + \varepsilon$$

$$H_1 : \underset{\sim}{\beta}'\underset{\sim}{\theta} \leq c \quad .$$

Suppose we know the prior distribution of $\underset{\sim}{\theta}$ is in the class $\Gamma = \{\tau |\ P_\tau[\underset{\sim}{\theta} \in H_0] = \lambda,\ P_\tau[\underset{\sim}{\theta} \in H_1] = \lambda'\}$, where $0 \leq \lambda, \lambda'$ and $\lambda + \lambda' \leq 1$. If the loss is defined as:

	a_0	a_1
H_0	0	L_1
H_1	L_2	0
$(H_0 \cup H_1)^c$	0	0

where a_0 means 'H_0 is true' and a_1 means 'H_1 is true'. To determine the Γ-minimax rule, we proceed as follows:

Solution: Let $\underset{\sim}{\beta}' = (\beta_1,\ldots,\beta_k)$, $\underset{\sim}{\theta}_0 = \frac{(c+\varepsilon)\underset{\sim}{\beta}}{\|\underset{\sim}{\beta}\|^2}$ and $\underset{\sim}{\theta}_1 = \frac{c\underset{\sim}{\beta}}{\|\underset{\sim}{\beta}\|^2}$. Let $\tau_0 \in \Gamma$ be such that $P_{\tau_0}[\underset{\sim}{\theta} = \underset{\sim}{\theta}_0] = \lambda$, and $P_{\tau_0}[\underset{\sim}{\theta} = \underset{\sim}{\theta}_1] = \lambda'$. Also, let $D = \{\delta \mid \delta$ is a measurable function on $\mathbb{R}^k$ such that $\delta(\underset{\sim}{x}) \in [0,1]\}$. For $\delta \in D$, $\delta(\underset{\sim}{x})$ is the probability of saying H_0 is true having observed $\underset{\sim}{X} = \underset{\sim}{x}$. Now,

$$r(\tau_0,\delta) = \int_X L_1\lambda(1-\delta(\underset{\sim}{x}))\, f(\underset{\sim}{x}-\underset{\sim}{\theta}_0) + L_2\lambda'\delta(\underset{\sim}{x})\, f(\underset{\sim}{x}-\underset{\sim}{\theta}_1)d\underset{\sim}{x},$$

so the Bayes rule wrt τ_0 is

$$\delta_0(\underset{\sim}{x}) = I_{[L_1\lambda f(\underset{\sim}{x}-\underset{\sim}{\theta}_0)\geq L_2\lambda' f(\underset{\sim}{x}-\underset{\sim}{\theta}_1)]}(\underset{\sim}{x}) .$$

But

$$L_1\lambda f(\underset{\sim}{x}-\underset{\sim}{\theta}_0) \geq L_2\lambda' f(\underset{\sim}{x}-\underset{\sim}{\theta}_1) \iff \underset{\sim}{x}'\underset{\sim}{\beta} \geq \frac{\|\underset{\sim}{\beta}\|^2}{\varepsilon}\,\ell n\,\frac{L_2\lambda' c(\underset{\sim}{\theta}_1)}{L_1\lambda\, c(\underset{\sim}{\theta}_0)} = k_0 .$$

Since $(\underset{\sim}{x} + t\underset{\sim}{\beta})'\underset{\sim}{\beta} \geq (\underset{\sim}{x} + s\underset{\sim}{\beta})'\underset{\sim}{\beta}$ if $t \geq s$, so $\delta_0(\underset{\sim}{x} + t\underset{\sim}{\beta}) \geq \delta_0(\underset{\sim}{x} + s\underset{\sim}{\beta})$, hence by Lemma 2.4.2,

$$E_{\underset{\sim}{\theta}+t\underset{\sim}{\beta}}[\delta_0(\underset{\sim}{X})] \geq E_{\underset{\sim}{\theta}+s\underset{\sim}{\beta}}[\delta_0(\underset{\sim}{X})] \quad \text{if} \quad t \geq s .$$

Now, let $\underset{\sim}{\theta} \in H_0$, then $\underset{\sim}{\theta} = v\underset{\sim}{\beta} + u\underset{\sim}{r}$ where $\underset{\sim}{\beta}'\underset{\sim}{r} = 0$. Since $\underset{\sim}{\beta}'\underset{\sim}{\theta} \geq c + \varepsilon \implies v \geq \frac{c+\varepsilon}{\|\underset{\sim}{\beta}\|^2}$, hence

$$E_{\underset{\sim}{\theta}}[\delta_0(\underset{\sim}{X})] \geq E_{\underset{\sim}{\theta}_0+u\underset{\sim}{r}}[\delta_0(\underset{\sim}{X})] .$$

But

$$\begin{aligned} E_{\underset{\sim}{\theta}_0+u\underset{\sim}{r}}[\delta_0(\underset{\sim}{X})] &= P[(\underset{\sim}{X}-\underset{\sim}{\theta}_0-u\underset{\sim}{r})'\underset{\sim}{\beta} \geq k_0 - \underset{\sim}{\theta}_0'\underset{\sim}{\beta} - u\underset{\sim}{r}'\underset{\sim}{\beta}] \\ &= P[\underset{\sim}{X}_0'\underset{\sim}{\beta} \geq k_0 - \underset{\sim}{\theta}_0'\underset{\sim}{\beta}] \quad \text{where} \quad \underset{\sim}{X}_0 \sim f(\underset{\sim}{x}) \\ &= E_{\underset{\sim}{\theta}_0}[\delta_0(\underset{\sim}{X})] , \end{aligned}$$

so we have proved that

$$\inf_{\underset{\sim}{\theta} \in H_0} E_{\underset{\sim}{\theta}}[\delta_0(\underset{\sim}{X})] = E_{\underset{\sim}{\theta}_0}[\delta_0(\underset{\sim}{X})] \ .$$

Similarly,

$$\sup_{\underset{\sim}{\theta} \in H_1} E_{\underset{\sim}{\theta}}[\delta_0(\underset{\sim}{X})] = E_{\underset{\sim}{\theta}_1}[\delta_0(\underset{\sim}{X})] \ .$$

Then we have for all $\tau \in \Gamma$,

$$\begin{aligned} r(\tau,\delta_0) &= \int_{H_0} L_1(1-E_{\underset{\sim}{\theta}}[\delta_0(\underset{\sim}{X})]\, d\tau(\underset{\sim}{\theta}) + \int_{H_1} L_2 E_{\underset{\sim}{\theta}}[\delta_0(\underset{\sim}{X})]\, d\tau(\underset{\sim}{\theta}) \\ &\le \lambda L_1(1-E_{\underset{\sim}{\theta}_0}[\delta_0(\underset{\sim}{X})]) + \lambda' L_2 E_{\underset{\sim}{\theta}_1}[\delta_0(\underset{\sim}{X})] \\ &= r(\tau_0,\delta_0) \ . \end{aligned}$$

Hence δ_0 is a Γ-minimax rule.

We have now displayed many examples for deriving the Γ-minimax rules (in Chapter 1: Γ-minimax rules for selecting populations close to a control; in Section 2.2 and 2.3 of this chapter: Γ-minimax rules to select the t-best populations). One might have noticed that the common setting of the Γ-minimax problems we have considered is that we have a partition on Θ such that $\Theta = (\bigcup_{i=1}^{r} \Theta_i) \cup \Theta_0$, where Θ_0 serves as an indifference zone, and Γ is defined as the class of prior distributions which put some known mass on each Θ_i $(1 \le i \le r)$. Then we will do the routine job of choosing a $\tau_0 \in \Gamma$. In general, τ_0 is a degenerate prior distribution which puts all the mass on the boundary of Θ_i. If we can prove that τ_0 is also the least favorable distribution in Γ for its Bayes rule δ_0, then δ_0 is a Γ-minimax rule.

All the problems that we have considered so far are under the assumption that all populations are independent, but this condition can be relaxed in certain problems. Let us look at the following example first:

Example 2.4.2. Let $\underset{\sim}{X} \sim N_k(\underset{\sim}{\theta},\Sigma)$ where $\Sigma = (1-\rho)\, I_k + \rho\, \underset{\sim}{1}_k \underset{\sim}{1}_k'$, ρ is known and $\rho > \frac{-1}{k-1}$. Let $\Theta = (\bigcup_{i=1}^{k} \Theta_i) \cup \Theta_0$ where $\Theta_i = \{\underset{\sim}{\theta} \mid \theta_i \geq \max_{j \neq i} \theta_j + \varepsilon\}$, $\Gamma = \{\tau \mid \int_{\Theta_i} d\tau(\underset{\sim}{\theta}) = \lambda_i\}$, with $\lambda_i \geq 0$ and $\sum_{i=1}^{k} \lambda_i \leq 1$ being given. Define

$$L(\underset{\sim}{\theta},\underset{\sim}{\delta}(\underset{\sim}{x})) = \begin{cases} 0 & \text{if } \theta \in \Theta_0 \\ L_i(1-\delta_i(\underset{\sim}{x})) + \ell_i \sum_{j \neq i} \delta_j(\underset{\sim}{x}) & \text{if } \theta \in \Theta_i, \end{cases}$$

where $\underset{\sim}{\delta} \in D = \{\underset{\sim}{\delta} \mid \sum_{i=1}^{k} \delta_i(\underset{\sim}{x}) = 1 \text{ and } \delta_i(x) \geq 0\}$.

We want to determine the Γ-minimax rule.

This problem is known as the selection of the best population and it was considered by Gupta and Huang (1977), for $\rho = 0$. When X_i's are equi-correlated, we let $\underset{\sim}{\theta}_i^* = (\theta_0,\ldots,\theta_0+\varepsilon,\ldots,\theta_0)$ for $1 \leq i \leq k$. Also, let τ_0 be the prior distribution such that $P_{\tau_0}[\underset{\sim}{\theta} = \underset{\sim}{\theta}_i^*] = \lambda_i$, then the Bayes rule wrt τ_0 is $\underset{\sim}{\delta}^0 = (\delta_1^0,\ldots,\delta_k^0)$ where

$$\delta_i^0(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } f_{\underset{\sim}{\theta}_i^*}(\underset{\sim}{x})\,(L_i+\ell_i)\lambda_i > \max_{j \neq i} f_{\underset{\sim}{\theta}_j^*}(L_j+\ell_j)\lambda j \\ r_j(\underset{\sim}{x}) & \qquad = \\ 0 & \qquad < \end{cases}$$

Now, let $\Sigma^{-1} = [\sigma^{ij}]$, then

$$\sigma^{ii} = \frac{1+(k-2)\rho}{[1+(k-1)\rho](1-\rho)} \qquad \text{for } 1 \le i \le k$$

$$\sigma^{ij} = \frac{-\rho}{[1+(k-1)\rho](1-\rho)} \qquad \text{for } i \neq j \quad .$$

then

$$f_{\underset{\sim}{\theta}_i^*}(\underset{\sim}{x})\ (L_i+\ell_i)\lambda_i > f_{\underset{\sim}{\theta}_j^*}(\underset{\sim}{x})\ (L_j+\ell_j)\lambda_j$$

$$\iff \frac{f_{\underset{\sim}{\theta}_i^*}(\underset{\sim}{x})}{f_{\underset{\sim}{\theta}_j^*}(\underset{\sim}{x})} > \frac{(L_j+\ell_j)\lambda_j}{(L_i+\ell_i)\lambda_i} = c_{ij}$$

$$\iff \varepsilon(\sigma^{ii}-\sigma^{ji})\ (x_i-x_j) > \ln c_{ij}$$

$$\iff x_i - x_j > \frac{1-\rho}{\varepsilon}\ \ln c_{ij} \quad .$$

We see that $\delta_i^0(\underset{\sim}{x})$ is increasing in x_i and is decreasing in x_j for $j \neq i$; also, $\delta_i^0(\underset{\sim}{x})$ is independent of the choice of θ_0. Hence we get

$$\inf_{\underset{\sim}{\theta}\in\Theta_i} E_{\underset{\sim}{\theta}}[\delta_i^0(\underset{\sim}{X})] = E_{(0,\ldots,0,\varepsilon,0,\ldots,0)}[\delta_i^0(\underset{\sim}{X})],$$

which proves that

$$r(\tau,\underset{\sim}{\delta}^0) = \sum_{i=1}^{k} \int_{\Theta_i} (\ell_i+L_i)\ (1-E_{\underset{\sim}{\theta}}[\delta_i^0(\underset{\sim}{X})])\ d\tau(\underset{\sim}{\theta}) \le r(\tau_0,\delta_0),$$

hence $\underset{\sim}{\delta}^0$ is a Γ-minimax rule.

The above example can be generalized to a more general theorem which we state as follows:

Theorem 2.4.1. Let $\underset{\sim}{X}$ has pdf as $f(\underset{\sim}{x}-\underset{\sim}{\theta})$. If the ratio

$$r_{ij}(\underset{\sim}{x}) = \frac{f(\underset{\sim}{x}-\varepsilon \underset{\sim}{e}_i)}{f(\underset{\sim}{x}-\varepsilon \underset{\sim}{e}_j)}$$

is an increasing function of x_i and is a decreasing function of x_j, keeping the other components fixed, then the problem of selecting the best population has a Γ-minimax rule $\underset{\sim}{\delta}^0 = (\delta_1^0,\ldots,\delta_k^0)$, where

$$\delta_\ell^0(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } r_{\ell j}(\underset{\sim}{x}) > c_{\ell j} \text{ for all } j \neq \ell \\ r_\ell(\underset{\sim}{x}) & \text{if } r_{\ell j}(\underset{\sim}{x}) \geq c_{\ell j} \text{ for all } j \neq \ell \text{ and '=' holds for some } j \neq \ell \\ 0 & \text{if } r_{\ell j}(\underset{\sim}{x}) < c_{\ell j} \text{ for some } j \end{cases}$$

Proof: Use the same argument as in Example 2.4.1 except the monotonicity of $r_{ij}(\underset{\sim}{x})$ is now guaranteed by the assumption instead of computation.

Remarks:

1. The monotonicity of $r_{ij}(\underset{\sim}{x})$ is satisfied if $f(\underset{\sim}{x}-\underset{\sim}{\theta}) = \prod_{i=1}^{k} g(x_i-\theta_i)$ and $g(x-\theta)$ has MLR in $\underset{\sim}{x}$.
2. If $Y_1,Y_2,\ldots Y_k,Z$ are $(k+1)$ independent random variables and $Y_i \sim g(y-\theta_i) = c(\theta_i)h(y)e^{\theta_i p(y)}$, where $p(y)$ is a strictly increasing function of y for $1 \leq i \leq k$, and Z is an arbitrary random variable with pdf as $g(z)$. Now, if $X_i = Y_i + Z$, then letting $\underset{\sim}{X} = (X_1,\ldots,X_k)$,

we have

$$\underset{\sim}{X} \sim f(\underset{\sim}{x}-\underset{\sim}{\theta}) = \int_{-\infty}^{\infty} \prod_{i=1}^{k} g(x_i - z - \theta_i) q(z)\, dz$$

and

$$r_{ij}(\underset{\sim}{x}) = \frac{f(\underset{\sim}{x}-\varepsilon \underset{\sim}{e}_i)}{f(\underset{\sim}{x}-\varepsilon \underset{\sim}{e}_j)}$$

$$= \frac{e^{\varepsilon p(x_i)} \int_{-\infty}^{\infty} \prod_{\ell=1}^{k} g(x_\ell - z) \frac{c(z+\varepsilon)}{c(z)} q(z)\, dz}{e^{\varepsilon p(x_j)} \int_{-\infty}^{\infty} \prod_{\ell=1}^{k} g(x_\ell - z) \frac{c(z+\varepsilon)}{c(z)} q(z)\, dz}$$

$$= e^{\varepsilon[p(x_i)-p(x_j)]} \quad .$$

Hence, the assumption of Theorem 2.3.1 is satisfied, and the Γ-minimax rule is

$$\delta_i^0(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } p(x_i) > \max_{j \neq i} p(x_j) + \frac{1}{\varepsilon} \ln c_{ij} \\ r_i(\underset{\sim}{x}) & \qquad = \\ 0 & \qquad < \end{cases}$$

Naturally, Example 2.4.1 is a special case in which $Y_i \sim N(\theta_i, 1-\rho)$ and $Z \sim N(0,\rho)$.

3. In Theorem 2.2.2, we assumed that X_i's are independent, because by independence, we can prove

$$x_i \ge x_j \Longrightarrow \sum_{s \in S_{i1} \setminus S_{j1}} f_{\underset{\sim}{\theta}_s^*}(\underset{\sim}{x}) \ge \sum_{s \in S_{j1} \setminus S_{i1}} f_{\underset{\sim}{\theta}_s^*}(\underset{\sim}{x}) \qquad (2.4.1)$$

$$\iff N_i(\underset{\sim}{x}) \le N_j(\underset{\sim}{x})$$

If X_i's are not independent but $X_i = Y_i + Z$ with Y_i and Z as defined in remark 2, one finds that when we choose $\theta_0 = 0$,

$$f_{\underset{\sim}{\theta}^*_S}(\underset{\sim}{x}) = e^{\varepsilon \sum_{i \in s(T)} p(x_i)} \int_{-\infty}^{\infty} \prod_{\ell=1}^{k} g(x_\ell - z)\left[\frac{c(z+\varepsilon)}{c(z)}\right]^t q(z)\, dz \; .$$

Hence (2.4.1) holds, and Theorem 2.2.2 is, therefore, still true.

As the last part of this section, we would like to search for the Γ-minimax rules for some hypothesis testing problems when $\underset{\sim}{X}$ has a multivariate density $f_{\underset{\sim}{\theta}}(\underset{\sim}{x})$, but $\underset{\sim}{\theta}$ is not a location parameter.

<u>Lemma 2.4.3.</u> Let $\underset{\sim}{X} = (X_1,\ldots,X_k)$ has pdf $f_{\underset{\sim}{\theta}}(\underset{\sim}{x})$. If the marginal distribution of $(X_2,\ldots,X_k)$ has pdf $g_{(\theta_2,\ldots,\theta_k)}(x_2,\ldots,x_k)$ and $X_1 \mid X_2,\ldots,X_k \sim h_{\eta(\underset{\sim}{\theta})}(x_1 \mid x_2,\ldots,x_k)$, where $\eta(\underset{\sim}{\theta})$ is an increasing function of θ_1 and $h_\eta(x_1 \mid x_2,\ldots,x_k)$ has MLR in x_1. Then if $\delta(\underset{\sim}{x})$ is an increasing function of x_1, we have that $E_{\underset{\sim}{\theta}}[\delta(\underset{\sim}{X})]$ is an increasing function of θ_1.

<u>Proof</u>: Let $\underset{\sim}{\theta} = (\theta_1,\theta_2,\ldots,\theta_k)$, $\underset{\sim}{\theta}' = (\theta_1',\theta_2,\ldots,\theta_k)$ with $\theta_1 \ge \theta_1'$, then $\eta(\underset{\sim}{\theta}) \ge \eta(\underset{\sim}{\theta}')$, so

$$E_{\eta(\underset{\sim}{\theta})}[\delta(\underset{\sim}{X}) \mid X_2,\ldots,X_k] \ge E_{\eta(\underset{\sim}{\theta}')}[\delta(\underset{\sim}{X}) \mid X_2,\ldots,X_k]$$

for all $X_2,\ldots,X_k$. Now,

$$E_{\underset{\sim}{\theta}}[\delta(\underset{\sim}{X})] = E_{\theta_2,\ldots,\theta_k}[E_{\eta(\underset{\sim}{\theta})}[\delta(\underset{\sim}{X}) \mid X_2,\ldots,X_k]]$$

$$\geq E_{\theta_2,\ldots,\theta_k}[E_{\eta(\underset{\sim}{\theta}')}[\delta(\underset{\sim}{X}) \mid X_2,\ldots,X_k]]$$

$$= E_{\underset{\sim}{\theta}'}[\delta(\underset{\sim}{X})] .$$

Examples of $f_{\underset{\sim}{\theta}}(\underset{\sim}{x})$ satisfying Lemma 2.4.3 are:

1. Multinomial distribution $MN(n,\underset{\sim}{\theta})$:

$$f_{\underset{\sim}{\theta}}(\underset{\sim}{x}) = \frac{n!}{\left(\prod_{i=1}^{k} x_i!\right)\left(n-\sum_{i=1}^{k} x_i\right)!}\left(\prod_{i=1}^{k}\theta_i^{x_i}\right)\left(1-\sum_{i=1}^{k}\theta_i\right)^{n-\sum_{i=1}^{k}x_i} \tag{2.4.2}$$

$$= \left[\frac{\left(n-\sum_{i=2}^{k}x_i\right)!}{x_1!\left(n-\sum_{i=1}^{k}x_i\right)!}\left(\frac{\theta_1}{1-\sum_{i=2}^{k}\theta_i}\right)^{x_1}\left(1-\frac{\theta_1}{1-\sum_{i=1}^{k}\theta_i}\right)^{n-\sum_{i=1}^{k}x_i}\right]$$

$$\cdot\left[\frac{n!}{\left(\prod_{i=1}^{k}x_i!\right)\left(n-\sum_{i=2}^{k}x_i\right)!}\prod_{i=2}^{k}\theta_i^{x_i}\left(1-\sum_{i=2}^{k}\theta_i\right)^{n-\sum_{i=2}^{k}x_i}\right]$$

We find that $\eta(\underset{\sim}{\theta}) = \dfrac{\theta_1}{1-\sum_{i=2}^{k}\theta_i}$ which is increasing in θ_1, and

$X_1 \mid X_2,\ldots,X_k \sim b(n-\sum_{i=2}^{k}x_i,\ \eta(\underset{\sim}{\theta}))$ which has the MLR, and the marginal

distribution of $(X_2,\ldots,X_k)' \sim MN(n,(\theta_2,\ldots,\theta_k)')$, so that Lemma 2.4.3 applies for this distribution. Furthermore, by the symmetry of the density of multinomial distribution, we get if $\delta(\underset{\sim}{x})$ is increasing (decreasing) in x_i, then $E_{\underset{\sim}{\theta}}(\delta(\underset{\sim}{X})]$ is increasing (decreasing) in θ_i for any $1 \le i \le k$.

2. Multivariate negative binomial distribution MNB$(n,\underset{\sim}{\theta})$:

$$f_{\underset{\sim}{\theta}}(\underset{\sim}{x}) = \frac{(n+\sum_{i=1}^{k} x_i-1)!}{(n-1)!\prod_{i=1}^{k}(x_i!)} \prod_{i=1}^{k} \theta_i^{x_i}\left(1+\sum_{i=1}^{k}\theta_i\right)^{-(n+\sum_{i=1}^{k}x_i)} \tag{2.4.3}$$

$$= \left[\frac{(n+\sum_{i=1}^{k} x_i-1)!}{(n+\sum_{i=1}^{k} x_i-1)!\, x_1!}\left(\frac{\theta_1}{1+\sum_{i=2}^{k}\theta_i}\right)^{x_1} \cdot \left(1+\frac{\theta_1}{1+\sum_{i=2}^{k}\theta_i}\right)^{-(n+\sum_{i=1}^{k}x_i)}\right]$$

$$\cdot\left[\frac{(n+\sum_{i=2}^{k} x_i-1)!}{(n-1)!\prod_{i=2}^{k} x_i!}\prod_{i=2}^{k}\theta_i^{x_i}\left(1+\sum_{i=2}^{k}\theta_i\right)^{-(n+\sum_{i=2}^{k}x_i)}\right]$$

We find $\eta(\underset{\sim}{\theta}) = \dfrac{\theta_1}{1+\sum_{i=2}^{k}\theta_i}$ which is increasing in θ_1,

$X_1 \mid X_2,\ldots,X_k \sim NB(n+\sum_{i=2}^{k}x_i,\ \eta(\underset{\sim}{\theta}))$, and

$(X_2,\ldots,X_k)' \sim MNB(n,(\theta_2,\ldots,\theta_k)')$, so the same result for multinomial distribution also holds for multivariate negative binomial distribution.

3. Multivariate normal distribution $N_k(\underset{\sim}{\theta},\Sigma)$:

Let

$$\Sigma = \begin{pmatrix} \sigma_{11} & \Sigma_{21}' \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad \underset{\sim}{\theta} = \begin{pmatrix} \theta_1 \\ \underset{\sim}{\theta}_2 \end{pmatrix}, \quad \text{and} \quad \underset{\sim}{X} = \begin{pmatrix} X_1 \\ \underset{\sim}{X}_2 \end{pmatrix},$$

then

$$\underset{\sim}{X}_2 \sim N_{k-1}(\underset{\sim}{\theta}_2,\Sigma_{22}) \quad \text{and} \quad X_1 \mid \underset{\sim}{X}_2 \sim N(\theta_1+\Sigma_{21}'\Sigma_{22}'(\underset{\sim}{X}_2-\underset{\sim}{\theta}_2),$$

$$\sigma_{11} - \Sigma_{21}'\Sigma_{22}^{-1}\Sigma_{22})$$

So Lemma 2.4.3 holds. But in the multivariate normal case, since $\underset{\sim}{\theta}$ is a location parameter, Lemma 2.4.2 is stronger than Lemma 2.4.3.

Example 2.4.3. Let $\underset{\sim}{X}_1,\underset{\sim}{X}_2,\ldots\underset{\sim}{X}_m$ be iid $MN(n,\underset{\sim}{\theta})$ with pdf as in (2.4.2). Let $\Theta = \{\underset{\sim}{\theta} \mid 0 < \theta_i$ for $1 \le i \le k$ and $\sum_{i=1}^{k} \theta_i < 1\}$. We want to test

$$H_0 : \sum_{j=1}^{t} \theta_j \ge a + \varepsilon$$

$$H_1 : \sum_{j=1}^{t} \theta_j \le a \quad ,$$

where $t \in \{1,2,\ldots,k\}$ and $0 < a < a + \varepsilon < 1$. If both Γ and the loss are the same as in Example 2.4.1, we can proceed as follows to get a Γ-minimax rule.

<u>Solution:</u> Let

$$\underset{\sim}{\theta}_0 = \left(\frac{a+\varepsilon}{t}, \dots, \frac{a+\varepsilon}{t}, \frac{1-a-\varepsilon}{k+1-t}, \dots, \frac{1-a-\varepsilon}{k+1-t} \right)$$

and

$$\underset{\sim}{\theta}_1 = \left(\frac{a}{t}, \dots, \frac{a}{t}, \frac{1-a}{k+1-t}, \dots, \frac{1-a}{k+1-t} \right) .$$

Let $\tau_0 \in \Gamma$ be such that $P_{\tau_0}[\underset{\sim}{\theta} = \underset{\sim}{\theta}_0] = \lambda$ and $P_{\tau_0}[\underset{\sim}{\theta} = \underset{\sim}{\theta}_1] = \lambda'$, then the Bayes rule wrt τ_0 is

$$\delta_0(\underset{\sim}{x}) = I_{[L_1 \lambda f_{\underset{\sim}{\theta}_0}(\underset{\sim}{x}) \geq L_2 \lambda' f_{\underset{\sim}{\theta}_1}(\underset{\sim}{x})]}(\underset{\sim}{x}).$$

Now,

$$L_1 \lambda f_{\underset{\sim}{\theta}_0}(\underset{\sim}{x}) \geq L_2 \lambda' f_{\underset{\sim}{\theta}_1}(\underset{\sim}{x})$$

$$\iff \prod_{j=1}^{t} \left(\frac{a+\varepsilon}{a} \cdot \frac{1-a}{1-a-\varepsilon} \right)^{x_j} \geq \frac{L_2 \lambda'}{L_1 \lambda} \left(\frac{1-a}{1-a-\varepsilon} \right)^n$$

$$\iff \sum_{j=1}^{t} x_j \geq \frac{\ln \frac{L_2 \lambda'}{L_1 \lambda} + n \ln \frac{1-a}{1-a-\varepsilon}}{\ln \frac{a+\varepsilon}{a} + \ln \frac{1-a}{1-a-\varepsilon}} = c_n .$$

Then, $\delta_0(\underset{\sim}{x}) = I_{[\sum_{j=1}^{t} x_j \geq c_n]}(\underset{\sim}{x})$ is increasing in x_j for all $1 \leq j \leq t$.

Hence, $E_{\underset{\sim}{\theta}}[\delta_0(\underset{\sim}{X})]$ is increasing in θ_j for all $1 \leq j \leq t$. Now, since $\sum_{j=1}^{t} x_j \sim b(n, \sum_{j=1}^{t} \theta_j)$, $E_{\underset{\sim}{\theta}}[\delta_0(\underset{\sim}{X})]$ depends only on $\sum_{j=1}^{t} \theta_j$. So we get

$$\inf_{\underset{\sim}{\theta} \in H_0} E_{\underset{\sim}{\theta}}[\delta_0(\underset{\sim}{X})] = E_{\underset{\sim}{\theta}_0}[\delta_0(\underset{\sim}{X})]$$

and

$$\sup_{\underset{\sim}{\theta} \in H_1} E_{\underset{\sim}{\theta}}[\delta_0(\underset{\sim}{X})] = E_{\underset{\sim}{\theta}_1}[\delta_0(\underset{\sim}{X})] .$$

It follows that $r(\tau,\delta_0) \le r(\tau_0,\delta_0)$ for all $\tau \in \Gamma$. If we consider $\sum_{i=1}^{m} \underset{\sim}{x}_i$ as the sufficient statistic for $\underset{\sim}{\theta}$ and $\sum_{i=1}^{m} \underset{\sim}{x}_i \sim MN(mn,\underset{\sim}{\theta})$, we get the Γ-minmax rule

$$\delta_0(\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_m) = I_{[\sum_{i=1}^{m} \sum_{j=1}^{t} x_{ij} \ge c_{mn}]}(\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_m) ,$$

where $\underset{\sim}{x}_i = (x_{i1},\ldots,x_{ik})$ for all $1 \le i \le m$.

<u>Example 2.4.4.</u> Let $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_m \sim MNB(n,\underset{\sim}{\theta})$ with density mass as in (2.4.3), $\Theta = \{\underset{\sim}{\theta} \mid 0 < \theta_i$ for all $1 \le i \le k\}$. We want to find Γ-minimax rule to test

$$H_0 : \sum_{j=1}^{t} \theta_j \ge a + \varepsilon$$

$$H_1 : \sum_{j=1}^{t} \theta_j \le a ,$$

where $a > 0$ and $t \in \{1,2,\ldots,k\}$. Γ and the loss, again, are the same as in Example 2.4.1.

<u>Solution</u>: Let $\underset{\sim}{\theta}_0 = \left(\frac{a+\varepsilon}{t} , \ldots , \frac{a+\varepsilon}{t} , \frac{1+a+\varepsilon}{k-t} , \ldots, \frac{1+a+\varepsilon}{k-t} \right)$

and

$$\underset{\sim}{\theta}_1 = \left(\frac{a}{t}, \ldots, \frac{a}{t} , \frac{1+a}{k-t} , \ldots , \frac{1+a}{k-t} \right) ,$$

then we find

$$\frac{f_{\theta_0}(\underset{\sim}{x})}{f_{\underset{\sim}{\theta}_0}(\underset{\sim}{x})} = \left(\frac{1+a+\varepsilon}{1+a}\right)^{-n} \prod_{j=1}^{t} \left[\frac{(a+\varepsilon)(1+a)}{a(1+a+\varepsilon)}\right]^{x_j} .$$

Hence, if τ_0 is such that $P_{\tau_0}[\underset{\sim}{\theta} = \underset{\sim}{\theta}_0] = \lambda$ and $P_{\tau_0}[\underset{\sim}{\theta} = \underset{\sim}{\theta}_1] = \lambda'$, the Bayes rule of τ_0 is

$$\delta_0(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \sum_{j=1}^{t} x_j \ge \dfrac{\ln \dfrac{L_2\lambda'}{L_1\lambda} + n \ln \dfrac{1+a+\varepsilon}{1+a}}{\ln \left[\dfrac{(a+\varepsilon)(1+a)}{a(1+a+\varepsilon)}\right]} = b_n \\ 0 & \qquad\quad < \end{cases}$$

Since $\sum_{j=1}^{t} x_j \sim NB(n, \sum_{j=1}^{t} \theta_j)$, so everything is the same as in Example 2.4.3, i.e., $\sup_{\tau\in\Gamma} r(\tau,\delta_0) = r(\tau_0,\delta_0)$. If we consider $\sum_{i=1}^{m} \underset{\sim}{x}_i$ as the sufficient statistic for $\underset{\sim}{\theta}$ and $\sum_{i=1}^{m} \underset{\sim}{x}_i \sim MNB(mn,\underset{\sim}{\theta})$, we get the Γ-minimax rule

$$\delta_0(\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_m) = I_{\left[\sum_{i=1}^{m} \sum_{j=1}^{t} x_{ij} \ge b_{mn}\right]} (\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_m) ,$$

where $\underset{\sim}{x}_i = (x_{i1},\ldots,x_{ik})$ for all $1 \le i \le m$.

CHAPTER III

EMPIRICAL BAYES RULES FOR SELECTING GOOD POPULATIONS

3.1. Introduction

We assume that G is an unknown prior distribution on Θ, and denote the minimum Bayes risk in a decision problem by $r(G)$. Robbins, in his pioneering papers (1955, 1964), proposed sequences of decision rules, based on data from n independent repetitions of the same decision problem, whose $(n+1)$st stage Bayes risk converges to $r(G)$ as $n \to \infty$. Such sequences of rules are called empirical Bayes rules. Empirical Bayes rules have been derived for multiple decision problems by Deely (1965), Van Ryzin (1970), Huang (1975), Van Ryzin and Susarla (1977), and Singh (1977). However, the forms of densities of the populations that these authors considered are either $c(\theta)h(x)e^{\theta x}$, for continuous case or $c(\theta)h(x)\theta^{x}$, for discrete case, and the loss functions are either squared error or merely $\max\limits_{1\leq j\leq k} \theta_j-\theta_i$ type. Fox (1978) discussed some estimation problem under squared error loss, in which empirical Bayes rule was derived for the first time for uniform distributions. Barr and Rizvi (1966), and McDonald (1974) also considered selection problems related to uniform distribution by the subset selection approach. It is interesting to note that uniform density is a good approximation to the central portion of normal density. We consider one

industrial example. One often wishes to keep the resistance in a circuit constant. If the resistance is normally distributed, then resistors with resistance in 1% tolerance interval about the mean are selected as "high quality". In this case, the uniform distribution can be used as a model for these high quality resistors. In Section 3.2, empirical Bayes rules are found for selecting populations better than a known control when the populations are uniformly distributed. In Section 3.3, the same problem is considered except that the control parameter is unknown. In Section 3.4, we derive the empirical Bayes rules for populations with densities of the form $p_i(x)c_i(\theta_i)I_{(0,\theta_i)}(x)$. Rate of convergence is also discussed in this section. Finally, Monte Carlo studies are carried out for the prior distribution $G(\theta) = \frac{\theta^2}{c^2} I_{(0,c)}(\theta)$. The smallest sample size N is determined to guarantee that the relative error is less than ε.

3.2. Known control parameter

Assume that we have k populations $\pi_1, \pi_2, \ldots, \pi_k$. $\pi_i \sim U(0,\theta_i)$ and θ_i is unknown for $1 \le i \le k$. Let θ_0 be a known control parameter, we define:

Definition 3.2.1. Population π_i is good iff $\theta_i > \theta_0$, and population π_i is bad iff $\theta_i \le \theta_0$.

Let $A = \{i | \theta_i > \theta_0\}$ and $B = \{i | \theta_i \le \theta_0\}$, then $A(B)$ is the set of indices of good (bad) populations. Our goal is to select good populations and reject bad ones. We formulate the problem in the empirical Bayes framework as follows:

(1) Let $\Theta = \{\underset{\sim}{\theta} = (\theta_1,\cdots,\theta_k) | \theta_i > 0$ for all $1 \le i \le k\}$ be the parameter space.

(2) Let $\mathcal{A} = \{S | S \subseteq \{1,2,\cdots,k\}\}$ be the action space. When we take action S, we say π_i is good if $i \in S$ and π_i is bad if $i \notin S$.

(3) Let $L: \Theta \times \mathcal{A} \to (0,\infty)$ be the loss function. We define

$$L(\underset{\sim}{\theta},S) = L_1 \sum_{i \in A \setminus S} (\theta_i - \theta_0) + L_2 \sum_{i \in B \cap S} (\theta_0 - \theta_i).$$

(4) Let $\underset{\sim}{G}(\underset{\sim}{\theta}) = \prod_{i=1}^{k} G_i(\theta_i)$ *be an unknown prior distribution on* Θ, where $G_i(\theta_i)$ has a continuous pdf $g_i(\theta_i)$.

(5) Let $(\Theta_{i1},Y_{i1}),\cdots,(\Theta_{in},Y_{in})$ be pairs of random variables from π_i and $Y_{ij}|\Theta_{ij} = \theta_{ij} \sim U(0,\theta_{ij})$ for all $1 \le i \le k$ and $1 \le j \le n$. Let $\underset{\sim}{Y}_j = (Y_{1j},\cdots,Y_{kj})$, then $\underset{\sim}{Y}_j$ denotes the previous j-th observations from $\pi_1,\cdots,\pi_k$.

(6) Let X_i be the present observation from π_i, for all $1 \le i \le k$. Let $\mathcal{X} = \{\underset{\sim}{x} = (x_1,\cdots,x_k) | x_i > 0$ for all $1 \le i \le k\}$. Also, let $\underset{\sim}{X} = (X_1,\cdots,X_k)$ and $f_{\underset{\sim}{\theta}}(\underset{\sim}{x}) = \prod_{i=1}^{k} \frac{1}{\theta_i} I_{(0,\theta_i)}(x_i)$. Since the loss function is bounded from below and we are interested in the Bayes rule, we can restrict our attention to the non-randomized rules. We have

(7) $D = \{\delta | \delta: \mathcal{X} \to \mathcal{A}$ *is a measurable function*$\}$. D is the collection of decision rules. Let

$$r(\underset{\sim}{G}) = \inf_{\delta \in D} r(\underset{\sim}{G},\delta) = r(\underset{\sim}{G},\delta^*), \qquad (3.2.1)$$

then δ^* is the Bayes rule wrt the prior distribution $\underset{\sim}{G}$, and $r(\underset{\sim}{G})$ is the minimum Bayes risk.

Definition 3.2.2. A sequence of decision rules $\{\delta_n(\underset{\sim}{x},\underset{\sim}{Y}_1,\cdots,\underset{\sim}{Y}_n)\}_{n=1}^{\infty}$ is said to be asymptotically optimal (a.o.) or empirical Bayes (e.B.) relative to G, if

$$r_n(\underset{\sim}{G},\delta_n) \equiv \int_{\mathcal{X}} [E\int_{\Theta} L(\theta,\delta_n(\underset{\sim}{x},\underset{\sim}{Y}_1,\cdots,\underset{\sim}{Y}_n))f_{\underset{\sim}{\theta}}(\underset{\sim}{x})d\underset{\sim}{G}(\underset{\sim}{\theta})]d\underset{\sim}{x}$$
$$\to r(\underset{\sim}{G}) \tag{3.2.2}$$

as $n \to \infty$. The expected value in (3.2.2) is taken wrt $\underset{\sim}{Y}_1,\cdots,\underset{\sim}{Y}_n$.

Remark: For simplicity, $\delta_n(\underset{\sim}{x},\underset{\sim}{Y}_1,\cdots,\underset{\sim}{Y}_n)$ will be denoted as $\delta_n(\underset{\sim}{x})$ from now on.

Let $m_i(x)$ be the marginal pdf of X_i and $M_i(x)$ be the marginal distribution of X_i. Then, we have

$$m_i(x) = \int_x^{\infty} \frac{1}{\theta_i}\, dG_i(\theta_i) \quad \text{for all } x > 0, \text{ and}$$

$$M_i(x) = \int_0^x\int_t^{\infty} \frac{1}{\theta}\, dG_i(\theta)dt$$
$$= \int_x^{\infty}\int_0^x \frac{1}{\theta}\, dtdG_i(\theta) + \int_0^x\int_0^{\theta} \frac{1}{\theta}\, dtdG_i(\theta)$$
$$= xm_i(x) + G_i(x).$$

Hence,

$$G_i(x) = M_i(x) - xm_i(x). \tag{3.2.3}$$

With the help of this formula, we are able to get a sequence of a.o. decision rules. As the first step, we would like to find $r(\underset{\sim}{G})$ and the associated Bayes rule. To get the Bayes rule easily, we will change the form of the loss function to the following:

$$L(\theta,S) = \sum_{i\in S}[L_2(\theta_0-\theta_i)I_{(0,\theta_0]}(\theta_i) - L_1(\theta_i-\theta_0)I_{(\theta_0,\infty)}(\theta_i)]$$
$$+ \sum_{i=1}^{k} L_1(\theta_i-\theta_0)I_{(\theta_0,\infty)}(\theta_i) \qquad (3.2.4)$$

It is easy to see that the second sum of (3.2.4) does not depend on the action S. Hence, to find the Bayes rule, we can omit the second sum and consider only the first sum in (3.2.4) as our loss function. Then,

$$r(\underset{\sim}{G},\delta) = \int_{\mathcal{X}} \sum_{i\in\delta(\underset{\sim}{x})} [\int_{\theta_i\leq\theta_0} L_2(\theta_0-\theta_i)f_{\underset{\sim}{\theta}}(\underset{\sim}{x})d\underset{\sim}{G}(\underset{\sim}{\theta})$$
$$- \int_{\theta_i>\theta_0} L_1(\theta_i-\theta_0)f_{\underset{\sim}{\theta}}(\underset{\sim}{x})d\underset{\sim}{G}(\underset{\sim}{\theta})]d\underset{\sim}{x}.$$

So, if $\delta^*(x) = S^*$ is the Bayes rule, then we find that $i \in S^*$ if

$$\int_{(0,\theta_0]\cap(x_i,\infty)} L_2(\theta_0-\theta_i)\frac{1}{\theta_i}\, dG_i(\theta_i) \leq \int_{\theta_0\vee x_i}^{\infty} L_1(\theta_i-\theta_0)\frac{1}{\theta_i}\, dG_i(\theta_i).$$

Hence, $i \in S^*$ if

(i) $x_i > \theta_0$, or

(ii) $x_i < \theta_0$ and

$$L_2\theta_0 \int_{x_i}^{\theta_0} \frac{1}{\theta_i}\, dG_i(\theta_i) - L_2[G_i(\theta_0)-G_i(x_i)]$$
$$\leq L_1(1-G_i(\theta_0)) - L_1\theta_0 \int_{\theta_0}^{\infty} \frac{1}{\theta_i}\, dG_i(\theta_i). \qquad (3.2.5)$$

The condition in (ii) is equivalent to $H_i(x_i) \le c_i(\theta_0)$, where

$$H_i(x_i) = L_2\theta_0 \int_{x_i}^{\theta_0} \frac{1}{\theta_i}\, dG_i(\theta_i) + L_2 G_i(x_i), \text{ and}$$

$$c_i(\theta_0) = L_2 G_i(\theta_0) + L_1(1-G_i(\theta_0)) - L_1\theta_0 \int_{\theta_0}^{\infty} \frac{1}{\theta_i}\, dG_i(\theta_i).$$

Since $H_i(x_i)$ is decreasing in x_i for $x_i < \theta_0$, so (i) and (ii) reduce to $x_i \ge \theta_0 - b_i$ where $b_i \ge 0$ and satisfies $H_i(b_i) = c_i(\theta_0)$. This shows that for any $\underset{\sim}{G}$, Gupta type rules are Bayes rules. Now, since $\underset{\sim}{G}$ is unknown, the Bayes rule is not obtainable. To find a.o. rules, we need to estimate $\underset{\sim}{G}$. In view of (3.2.3), we need to estimate M_i and m_i.

Let $\{Y_i\}_{i=1}^{\infty}$ be a sequence of iid random variables with a common distribution function $K(y)$. We also assume that $K'(y) = k(y)$ exists a.e.. Let

$$K_n(y) = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty,y]}(Y_i) \tag{3.2.6}$$

and

$$k_n(y) = \frac{1}{h}[K_n(y+h) - K_n(y)]. \tag{3.2.7}$$

Then, $K_n(y) \to K(y)$ uniformly in y with probability 1 (Glivenko-Cantelli Theorem) as $n \to \infty$. The following lemma guarantees the convergence of $k_n(y)$ to $k(y)$.

Lemma 3.2.1. (Parzen (1962))

(i) If $h = h(n)$ in (3.2.7) satisfies $\lim_{n\to\infty} h(n) = 0$, then

$\lim_{n\to\infty} E[k_n(y)] = k(y)$ for any continuous point y of $k(\cdot)$.

(ii) If in addition to (i), $h(n)$ also satisfies $\lim_{n\to\infty} nh(n) = \infty$,

then $\lim_{n\to\infty} E|k_n(y)-k(y)|^2 = 0$ for any continuous point y of

$k(\cdot)$.

Remarks:

1. In our problem $m_i(y) = \int_y^\infty \frac{1}{\theta}\, dG_i(\theta)$, hence $m_i(y)$ is continuous at all y. So, (i) and (ii) in Lemma 3.2.1 hold for all y.

2. By Chebyshev's inequality and (ii), we have

$$\lim_{n\to\infty} P[|k_n(y)-k(y)| > \varepsilon] \le \lim_{n\to\infty} \frac{E|k_n(y)-k(y)|^2}{\varepsilon^2} = 0.$$

Hence, if $h \to 0$ and $nh \to 0$, it is shown that $k_n(y) \to k(y)$ in (p).

Now, we state a theorem which provides a sufficient condition for $\{\delta_n(\underset{\sim}{x})\}_{n=1}^\infty$ to be empirical Bayes. Let

$$\Delta_{G_i}(x_i) = H_i(x_i) - c_i(\theta_0) \tag{3.2.8}$$

and

$$S_0(\underset{\sim}{x}) = \{i \mid x_i < \theta_0 \text{ and } \Delta_{G_i}(x_i) \le 0\}.$$

Now, for any i $(1 \leq i \leq k)$, let $\Delta_{i,n}(x_i) = \Delta_i(x_i, Y_{i1}, \cdots, Y_{in})$ for all $n = 1,2,\cdots$, be a sequence of measurable real-valued functions, we define

$$S_n(\underset{\sim}{x}) = \{i \mid x_i < \theta_0 \text{ and } \Delta_{i,n}(x_i) \leq 0\} \tag{3.2.9}$$

and

$$\delta_n^*(\underset{\sim}{x}) = \{i \mid x_i \geq \theta_0\} \cup S_n(\underset{\sim}{x}) \tag{3.2.10}$$

Then we claim

Theorem 3.2.1. If for $1 \leq i \leq k$, $\int_0^\infty \theta_i dG_i(\theta_i) < \infty$ and $\Delta_{i,n}(x_i) \to \Delta_{G_i}(x_i)$ in (p) for almost all $x_i < \theta_0$. Then $\{\delta_n^*(\underset{\sim}{x})\}_{n=1}^\infty$ defined by (3.2.10) is empirical Bayes.

Proof: For all $S \in \mathscr{A}$, let

$$\mathscr{X}_S = \{\underset{\sim}{x} \in \mathscr{X} \mid x_i \geq \theta_0 \text{ if } i \in S \text{ and } x_i < \theta_0 \text{ if } i \notin S\}.$$

Now, for any $\underset{\sim}{x} \in \mathscr{X}_S$, $\delta^*(\underset{\sim}{x}) = S \cup S_0(\underset{\sim}{x})$. Hence, for $\underset{\sim}{x} \in \mathscr{X}_S$,

$$\int_{\Theta} L(\theta, \delta^*(\underset{\sim}{x})) f_{\underset{\sim}{\theta}}(\underset{\sim}{x}) d\underset{\sim}{G}(\underset{\sim}{\theta})$$

$$= \sum_{i \in \delta^*(\underset{\sim}{x})} \left[\int_{\theta_i \leq \theta_0} L_2(\theta_0 - \theta_i) f_{\underset{\sim}{\theta}}(\underset{\sim}{x}) d\underset{\sim}{G}(\underset{\sim}{\theta}) - \int_{\theta_i > \theta_0} L_1(\theta_i - \theta_0) f_{\underset{\sim}{\theta}}(\underset{\sim}{x}) d\underset{\sim}{G}(\underset{\sim}{\theta})\right]$$

$$= \sum_{i \in S} \left[-\int_{\theta_i > \theta_0} L_1(\theta_i - \theta_0) f_{\underset{\sim}{\theta}}(\underset{\sim}{x}) d\underset{\sim}{G}(\underset{\sim}{\theta})\right] + \sum_{i \in S_0(\underset{\sim}{x})} \Delta_{G_i}(x_i) \prod_{\substack{j=1 \\ j \neq i}}^{k} m_j(x_j).$$

Similarly, for $\underset{\sim}{x} \in \mathscr{X}_S$, we have

$$\int_{\Theta} L(\underset{\sim}{\theta},\delta_n^*(\underset{\sim}{x}))f_{\underset{\sim}{\theta}}(\underset{\sim}{x})d\underset{\sim}{G}(\underset{\sim}{\theta})$$

$$= \sum_{i\in S}[-\int_{\theta_i>\theta_0} L_1(\theta_i-\theta_0)f_{\underset{\sim}{\theta}}(\underset{\sim}{x})d\underset{\sim}{G}(\underset{\sim}{\theta})] + \sum_{i\in S_n(\underset{\sim}{x})} \Delta_{G_i}(x_i)\prod_{\substack{j=1\\j\neq i}}^{k} m_j(x_j).$$

Hence, if

$$\Delta_{i,n}(x_i) \to \Delta_{G_i}(x_i) \quad \text{in} \quad (p), \quad \text{then}$$

$$0 \le \int_{\Theta} L(\underset{\sim}{\theta},\delta_n^*(\underset{\sim}{x}))f_{\underset{\sim}{\theta}}(\underset{\sim}{x})d\underset{\sim}{G}(\underset{\sim}{\theta}) - \int_{\Theta} L(\underset{\sim}{\theta},\delta^*(\underset{\sim}{x}))f_{\underset{\sim}{\theta}}(\underset{\sim}{x})d\underset{\sim}{G}(\underset{\sim}{\theta})$$

$$\le \sum_{i\in S_n(\underset{\sim}{x})} |\Delta_{G_i}(x_i)-\Delta_{i,n}(x_i)| \prod_{\substack{j=1\\j\neq i}}^{k} m_j(x_j)$$

$$+\Big(\sum_{i\in S_n(\underset{\sim}{x})} - \sum_{i\in S_0(\underset{\sim}{x})}\Big)\Delta_{i,n}(x_i) \prod_{\substack{j=1\\j\neq i}}^{k} m_j(x_j) \tag{3.2.11}$$

$$+ \sum_{i\in S_0(\underset{\sim}{x})} |\Delta_{i,n}(x_i)-\Delta_{G_i}(x_i)| \prod_{\substack{j=1\\j\neq i}}^{k} m_j(x_j)$$

$$\le 2\varepsilon \sum_{i=1}^{k} \prod_{\substack{j=1\\j\neq i}}^{k} m_j(x_j)$$

with probability near 1 for all $n > N$. Note that (3.2.11) is non-positive by the definition of $S_n(\underset{\sim}{x})$. Now, we have proved that

$$\int_{\Theta} L(\underset{\sim}{\theta},\delta_n^*(\underset{\sim}{x}))f_{\underset{\sim}{\theta}}(\underset{\sim}{x})d\underset{\sim}{G}(\underset{\sim}{\theta}) \to \int_{\Theta} L(\underset{\sim}{\theta},\delta^*(\underset{\sim}{x}))f_{\underset{\sim}{\theta}}(\underset{\sim}{x})d\underset{\sim}{G}(\underset{\sim}{\theta})$$

in (p), for almost all $\underset{\sim}{x}$. By Corollary 1 of Robbins (1964), we conclude that $\{\delta_n^*(\underset{\sim}{x},\underset{\sim}{Y}_1,\ldots,\underset{\sim}{Y}_n)\}_{n=1}^{\infty}$ is empirical Bayes. Thus completes the proof.

We have reduced the problem of finding the empirical Bayes rules to the problem of finding a consistent estimator of $\Delta_{G_i}(x_i)$ in (p). If we recall that

$$m_i(x_i) = \int_{x_i}^{\infty} \frac{1}{\theta_i}\, dG_i(\theta_i) \quad \text{and} \quad G_i(x) = M_i(x) - x m_i(x),$$

Then from (3.2.8) we get

$$\Delta_{G_i}(x_i) = L_2 m_i(x_i)(\theta_0 - x_i) + L_2[M_i(x_i) - M_i(\theta_0)] + L_1[M_i(\theta_0) - 1].$$

Hence, if we define

$$\Delta_{i,n}(x_i) = L_2 m_{i,n}(x_i)(\theta_i - x_i) + L_2[M_{i,n}(x_i) - M_{i,n}(\theta_0)]$$
$$+ L_1[M_{i,n}(\theta_0) - 1], \tag{3.2.12}$$

where

$$M_{i,n}(x) = \frac{1}{n} \sum_{j=1}^{n} I_{(0,x]}(Y_{ij}) \tag{3.2.13}$$

and

$$m_{i,n}(x) = \frac{1}{h}[M_{i,n}(x+h) - M_{i,n}(x)] = \frac{1}{nh} \sum_{j=1}^{n} I_{(x,x+h]}(Y_{ij}) \tag{3.2.14}$$

then by Lemma 3.2.1,

$$\Delta_{i,n}(x_i) \to \Delta_{G_i}(x_i) \quad \text{in} \quad (p)$$

for all $\underset{\sim}{x}$. Thus, the sequence of rules $\{\delta_n^*(\underset{\sim}{x})\}_{n=1}^{\infty}$ which is defined by (3.2.10), with $\Delta_{i,n}(x_i)$ defined by (3.2.12), is empirical Bayes.

3.3. θ_0 unknown

In this section, Π_0 is a control population which is distributed as $U(0,\theta_0)$ with θ_0 unknown. Let $Y_{01},\cdots,Y_{0n}$ be the past data collected from Π_0, and let X_0 be the present observation from Π_0. Based on this further information, we will search for empirical Bayes rules for selecting populations better than the control. Note that now $\underset{\sim}{\theta} = (\theta_0,\theta_1,\cdots,\theta_k)$, $\underset{\sim}{x} = (x_0,x_1,\cdots,x_k)$ and $\underset{\sim}{G}(\underset{\sim}{\theta}) = \prod_{i=0}^{k} G_i(\theta_i)$. When the same loss function is used, the Bayes rule δ^* now becomes: $i \in \delta^*(\underset{\sim}{x})$ if

$$L_2 \int_{x_0}^{\infty} \frac{1}{\theta_0} \int_{(0,\theta_0]\cap(x_i,\infty)} \frac{1}{\theta_i} (\theta_0-\theta_i)dG_i(\theta_i)dG_0(\theta_0)$$

$$\leq L_1 \int_{x_0}^{\infty} \frac{1}{\theta_0} \int_{(\theta_0,\infty)\cap(x_i,\infty)} \frac{1}{\theta_i} (\theta_i-\theta_0)dG_i(\theta_i)dG_0(\theta_0).$$

Hence, $i \in \delta^*(\underset{\sim}{x})$ if

(i) $x_i \geq x_0$ and $\Delta^1_{G_0G_i}(x_0,x_i) \leq 0$, where

$$\Delta^1_{G_0,G_i}(x_0,x_i) = (L_1-L_2)\left[\int_{x_i}^{\infty} m_i(\theta_0)dG_0(\theta_0)+\int_{x_i}^{\infty} m_0(\theta_i)dG_i(\theta_i)\right]$$

$$- L_1[1-G_i(x_i)]m_0(x_0)+m_i(x_i)[L_2+(L_1-L_2)G_0(x_i)-L_1G_0(x_0)], \tag{3.3.1}$$

(ii) $x_i < x_0$ and $\Delta^2_{G_0G_i}(x_0,x_i) \leq 0$, where

$$\Delta^2_{G_0G_i}(x_0,x_i) = (L_1-L_2)[\int_{x_0}^{\infty} m_i(\theta_0)dG_0(\theta_0)+\int_{x_0}^{\infty} m_0(\theta_i)dG_i(\theta_i)]$$

$$- m_0(x_0)[L_1+(L_2-L_1)G_i(x_0)-L_2G_i(x_i)]+L_2m_i(x_i)(1-G_0(x_0)). \tag{3.3.2}$$

When $L_1 = L_2 = L$, the Bayes rule is greatly simplified. Then we have $i \in S^*(\underset{\sim}{x})$ if

$$\Delta_{G_0,G_i}(x_0,x_i) = m_0(x_0)[1-G_i(x_i)]-m_i(x_i)[1-G_0(x_0)] \geq 0.$$

Now, a consistent estimator of $\Delta_{G_0,G_i}(x_0,x_i)$ is obtained by

$$\Delta_{i,n}(x_i,x_0) = m_{0,n}(x_0)[1-G_{i,n}(x_i)]-m_{i,n}(x_i)[1-G_{0,n}(x_0)]$$

where $m_{i,n}(x)$ is defined by (3.2.13) and (3.2.14), and $G_{i,n}(x) = M_{i,n}(x) - xm_{i,n}(x)$ for all $0 \leq i \leq n$. Let

$\delta^*_n(\underset{\sim}{x}) = \{i|\Delta_{i,n}(x_i,x_0) \geq 0\}$, then

$\{\delta^*_n(\underset{\sim}{x})\}_{n=1}^{\infty}$ are empirical Bayes by Theorem 3.3.2.

When $L_1 \neq L_2$, we need to find consistent estimators of

$$\int_a^{\infty} m_i(\theta_0)dG_0(\theta_0) \quad \text{and} \quad \int_a^{\infty} m_0(\theta_i)dG_i(\theta_i).$$

The next theorem provides us with such estimators.

Theorem 3.3.1. Let $M_{i,n}(x)$ and $m_{i,n}(x)$ be defined by (3.2.13) and (3.2.14), respectively, for all $1 \leq i \leq k$. If $h > 0$, $h \to 0$, and $nh^2 \to \infty$ as $n \to \infty$, and if

$\int_0^\infty \theta_i dG_i(\theta_i) < \infty$ for all $0 \le i \le k$, then

$$-\int_a^\infty x m_{i,n}(x) dm_{0,n}(x) \to \int_a^\infty m_i(x) dG_0(x) \text{ in } (p)$$

for any $a > 0$.

Proof: $\int_a^\infty x m_{i,n}(x) dm_{0,n}(x)$

$$= \frac{1}{n^2}\frac{1}{h^2}\sum_{j=1}^{n}\sum_{\ell=1}^{n}\int_a^\infty x I_{(x,x+h]}(Y_{ij}) dI_{[Y_{0\ell}-h,Y_{0\ell})}(x)$$

$$= \frac{1}{n^2}\frac{1}{h^2}\sum_{j=1}^{n}\sum_{\ell=1}^{n}(U_{j\ell}-V_{j\ell}), \text{ where}$$

$$U_{j\ell} = (Y_{0\ell}-h) I_{(a,\infty)}(Y_{0\ell}-h) I_{(Y_{0\ell}-h,Y_{0\ell}]}(Y_{ij}), \text{ and}$$

$$V_{j\ell} = Y_{0\ell} I_{(a,\infty)}(Y_{0\ell}) I_{(Y_{0\ell},Y_{0\ell}+h]}(Y_{ij}).$$

Now, since $Y_{0\ell} \sim M_0(x)$ and $Y_{ij} \sim M_i(x)$ for all $1 \le j$, $\ell \le n$, we have

$$E\int_a^\infty x m_{i,n}(x) dm_{0,n}(x)$$

$$= \int_a^\infty x \frac{1}{h}\int_x^{x+h} dM_i(y)\,\frac{1}{h}[m_0(x+h)-m_0(x)]dx. \text{ Also, because}$$

$$\frac{1}{h}\int_x^{x+h} dM_i(y) = \frac{1}{h}\int_x^{x+h}\int_y^\infty \frac{1}{\theta}\, dG_i(\theta)dy$$

$$\le \frac{1}{h}\int_x^{x+h} dy \int_x^\infty \frac{1}{\theta}\, dG_i(\theta) \le \frac{1}{x}(1-G_i(x)) \qquad (3.3.3)$$

so,

$$|x \frac{1}{h} \int_x^{x+h} dM_i(y) \frac{1}{h} [m_0(x+h)-m_0(x)]| \le \frac{1}{h} \int_x^{x+h} \frac{1}{\theta} dG_0(\theta),$$

hence by LDCT, we get

$$\lim_{n\to\infty} E \int_a^\infty xm_{i,n}(x)dm_{0,n}(x) = \int_a^\infty xm_i(x)m_0'(x)dx$$

$$= - \int_a^\infty m_i(x)dG_0(x). \tag{3.3.4}$$

Now,

$$\mathrm{Var} \int_a^\infty xm_i^n(x)dm_0^n(x) = \mathrm{Var} \frac{1}{n^2} \frac{1}{h^2} \sum_{j,\ell} (U_{j\ell}-V_{j\ell})$$

$$= \frac{1}{n^4h^4} \{ \sum_{j,\ell} \mathrm{Var}(U_{j\ell}-V_{j\ell}) + \sum_{j=1}^{n} \sum_{\ell_1\neq\ell_2} \mathrm{Cov}(U_{j\ell_1}-V_{j\ell_1}, U_{j\ell_2}-V_{j\ell_2})$$

$$+ \sum_{\ell=1}^{n} \sum_{j_1\neq j_2} \mathrm{Cov}(U_{j_1\ell}-V_{j_1\ell}, U_{j_2\ell}-V_{j_2\ell}) \}$$

$$= \frac{1}{n^2h^4} \mathrm{Var}(U_{11}-V_{11}) + \frac{n-1}{n^2h^4} \mathrm{Cov}(U_{11}-V_{11}, U_{12}-V_{12})$$

$$+ \frac{n-1}{n^2h^4} \mathrm{Cov}(U_{11}-V_{11}, U_{21}-V_{21}), \tag{3.3.5}$$

but $\mathrm{Var}(U_{11}-V_{11}) \le E[(U_{11}-V_{11})^2] = E(U_{11}^2) + E(V_{11}^2)$, and

$$\frac{1}{h} E(U_{11}^2) = \int_a^\infty x^2 \frac{1}{h} \int_x^{x+h} dM_i(y)dM_0(x+h)$$

$$\le \int_a^\infty x^2 \frac{1}{x} [1-G_i(x)] \int_{x+h}^\infty \frac{1}{\theta} dG_0(\theta)$$

$$\leq \int_a^\infty [1-G_i(x)]dx$$

$$\leq \int_0^\infty \theta dG_i(\theta), \qquad \text{by (3.3.3)}$$

so $\frac{1}{h} E(U_{11}^2)$ is bounded for all h. Similarly, we can prove that $\frac{1}{h} E(V_{11}^2) \leq \int_0^\infty \theta dG_0(\theta)$, hence

$$\frac{1}{h} \operatorname{Var}(U_{11}-V_{11}) \leq 2 \int_0^\infty \theta dG_0(\theta). \tag{3.3.6}$$

Meanwhile,

$$\operatorname{Cov}(U_{11}-V_{11},U_{12}-V_{12}) = \operatorname{Cov}(U_{11},U_{12}) + \operatorname{Cov}(V_{11},V_{12})$$

$$- \operatorname{Cov}(U_{11},V_{12}) - \operatorname{Cov}(V_{11},U_{12}),$$

and

$$\left|\frac{1}{h^2} \operatorname{Cov}(U_{11},U_{12})\right| \leq \frac{1}{h^2} \{E(U_{11},U_{12})+E(U_{11})E(U_{12})\} \leq 2(1-M_i(a)).$$

Similarly, we can prove that $2(1-M_i(a))$ is also an upper bound for $\left|\frac{1}{h^2} \operatorname{Cov}(V_{11},V_{12})\right|$, $\left|\frac{1}{h^2} \operatorname{Cov}(U_{11},V_{12})\right|$ and $\left|\frac{1}{h^2} \operatorname{Cov}(V_{11},U_{12})\right|$. Hence

$$\left|\frac{1}{h^2} \operatorname{Cov}(U_{11}-V_{11},U_{12}-V_{12})\right| \leq 8(1-M_i(a)). \tag{3.3.7}$$

Finally, $\left|\frac{1}{h^2} \operatorname{Cov}(U_{11}-V_{11},U_{21}-V_{21})\right| \leq 8(1-M_i(a)).$ (3.3.8)

Hence, from (3.3.5) we get

$$\operatorname{Var} \int_a^\infty xm_{i,n}(x)dm_{0,n}(x) \to 0$$

if $nh^2 \to \infty$ and $h \to 0$ by (3.3.6), (3.3.7) and (3.3.8). The fact that the variance goes to 0 and the expected value converges to $-\int_a^\infty m_i(x)dG_0(x)$ as shown in (3.3.4) implies that

$$\int_a^\infty x m_{i,n}(x)dm_{0,n}(x) \to -\int_a^\infty m_i(x)dG_0(x) \quad \text{in} \quad (p)$$

by Chebyshev's inequality. This finishes the proof.

Recall that the Bayes rule is

$$\delta^*(\underset{\sim}{x}) = \{i | x_i \geq x_0 \text{ and } \Delta^1_{G_0G_i}(x_0,x_i) \leq 0\}$$

$$\cup \{i | x_i < x_0 \text{ and } \Delta^2_{G_0G_i}(x_0x_i) \leq 0\}$$

$$\equiv S_1^*(\underset{\sim}{x}) \cup S_2^*(\underset{\sim}{x}),$$

where $\Delta^1_{G_0,G_i}(x_0,x_i)$ and $\Delta^2_{G_0,G_i}(x_0,x_i)$ are defined by (3.3.1) and (3.3.2) respectively. Now, Theorem 3.2.1 has a similar version for θ_0 unknown.

<u>Theorem 3.3.2</u>. Assume that $\int_0^\infty \theta dG_i(\theta) < \infty$ for all $0 \leq i \leq k$. If for all $1 \leq i \leq k$, $\Delta^1_{i,n}(x_0,x_i) \to \Delta^1_{G_i,G_0}(x_0,x_i)$ in (p) for $x_i \geq x_0$, and $\Delta^2_{i,n}(x_0,x_i) \to \Delta^2_{G_i,G_0}(x_0,x_i)$ in (p) for $x_i < x_0$. Then if we let

$$\delta_n^*(\underset{\sim}{x}) = \{i | x_i \geq x_0 \text{ and } \Delta^1_{i,n}(x_0,x_i) \leq 0\}$$

$$\cup \{i | x_i < x_0 \text{ and } \Delta^2_{i,n}(x_0,x_i) \leq 0\}$$

$$= S_n^1(\underset{\sim}{x}) \cup S_n^2(\underset{\sim}{x}),$$

we have $\{\delta_n^*(\underset{\sim}{x})\}_{n=1}^{\infty}$ is empirical Bayes.

<u>Proof</u>:

$$\int_{\Theta} L(\underset{\sim}{\theta},\delta^*(\underset{\sim}{x}))f_{\underset{\sim}{\theta}}(\underset{\sim}{x})d\underset{\sim}{G}(\underset{\sim}{\theta})$$

$$= \sum_{i\in S_1^*(\underset{\sim}{x})} \Delta^1_{G_i,G_0}(x_0,x_i) \prod_{\substack{j=1\\ j\neq i}}^{k} m_j(x_j) + \sum_{i\in S_2^*(\underset{\sim}{x})} \Delta^2_{G_i,G_0}(x_0,x_i) \prod_{\substack{j=1\\ j\neq i}}^{k} m_j(x_j)$$

and

$$\int_{\Theta} L(\underset{\sim}{\theta},\delta_n^*(\underset{\sim}{x}))f_{\underset{\sim}{\theta}}(\underset{\sim}{x})d\underset{\sim}{G}(\underset{\sim}{\theta})$$

$$= \sum_{i\in S_n^1(\underset{\sim}{x})} \Delta^1_{G_i,G_0}(x_0,x_i) \prod_{\substack{j=1\\ j\neq i}}^{k} m_j(x_j) + \sum_{i\in S_n^2(\underset{\sim}{x})} \Delta^2_{G_i,G_0}(x_0,x_i) \prod_{\substack{j=1\\ j\neq i}}^{k} m_j(x_j).$$

Now, following the same method as in the proof of Theorem 3.2.1, we can show

$$\sum_{i\in S_n^{\ell}(\underset{\sim}{x})} \Delta^{\ell}_{G_i,G_0}(x_i,x_0) \prod_{\substack{j=1\\ j\neq i}}^{k} m_j(x_j) \to \sum_{i\in S_{\ell}^*(\underset{\sim}{x})} \Delta^{\ell}_{G_i,G_0}(x_i,x_0) \prod_{\substack{j=1\\ j\neq i}}^{k} m_j(x_j)$$

in (p) for $\ell = 1,2$. Hence by Lemma 3.2.2, $\{\delta_n^*(\underset{\sim}{x})\}_{n=1}^{\infty}$ is empirical Bayes. This completes the proof.

Now, let

$$\Delta^1_{i,n}(x_i,x_0) = (L_2-L_1)\Big[\int_{x_i}^{\infty} xm_{i,n}(x)dm_{0,n}(x)$$

$$+ \int_{x_i}^{\infty} xm_{0,n}(x)dm_{i,n}(x)\Big]-L_1[1-G_{i,n}(x_i)]m_{0,n}(x_0)$$

$$+ m_{i,n}(x_i)[L_2+(L_1-L_2)G_{0,n}(x_i)-L_1G_{0,n}(x_0)] \qquad (3.3.10)$$

and

$$\Delta^2_{i,n}(x_i,x_0) = (L_2-L_1)[\int_{x_0}^{\infty} xm_{i,n}(x)dm_{0,n}(x)$$

$$+ \int_{x_0}^{\infty} xm_{0,n}(x)dm_{i,n}(x)]+L_2[1-G_{0,n}(x_0)]m_{i,n}(x_i)$$

$$- m_{0,n}(x_0)[L_1+(L_2-L_1)G_{i,n}(x_0)-L_2G_{i,n}(x_i)], \tag{3.3.11}$$

then by Theorem 3.3.1, the conditions of Theorem 3.3.2 are satisfied. Hence, (3.3.9), (3.3.10) and (3.3.11) define a sequence of empirical Bayes rules.

3.4. Generalization and simulation

Let $p_i(x)$ be a positive continuously differentiable function which is defined over $(0,\infty)$ for all $1 \leq i \leq k$. Also let $c_i(\theta)^{-1} = \int_0^{\theta} p_i(x)dx$ for $\theta > 0$, then $f_i(x|\theta) = p_i(x)c_i(\theta)I_{(0,\theta)}(x)$ is a density function. In this section, we assume that population $\Pi_i \sim f_i(x|\theta_i)$ for all $1 \leq i \leq k$. Under the formulation of Section 3.2, we wish to find the empirical Bayes rules for these more general density functions. For simplicity, we assume that $L_1 = L_2 = L$ and that θ_0 is known. Also, we assume that $G_i(\theta)$ has a continuous density $g_i(\theta)$ with a bounded support $[0,\alpha_i]$, and α_i is known for all $1 \leq i \leq k$. Now,

$$m_i(x) = \int_0^{\infty} f_i(x|\theta)dG_i(\theta) = p_i(x)\int_x^{\alpha_i} c_i(\theta)dG_i(\theta).$$

If we follow the same discussion as in Section 3.2, we can show that the Bayes rule δ^* is: $i \in \delta^*(\underset{\sim}{x})$ iff

(i) $x_i \geq \theta_0$, or

(ii) $x_i < \theta_0$ and $\theta_0 \int_{x_i}^{\alpha_i} c_i(x)dG_i(x) \leq \int_{x_i}^{\alpha_i} xc_i(x)dG_i(x)$.

Hence, we find $i \in \delta^*(\underset{\sim}{x})$ iff $x_i \geq \theta_0 - d_i$ where d_i satisfies $\int_{d_i}^{\alpha_i}(\theta_0 - x)c_i(x)dG_i(x) = 0$. Let $d_{i,n} = d_{i,n}(Y_{i1},\cdots,Y_{in})$ be a consistent estimation of d_i, then $\delta_n^0(\underset{\sim}{x}) = \{i|x_i \geq \theta_0 - d_{i,n}\}$ defines a sequence of empirical Bayes rules, and these are (weak) admissible in the sense that $\delta_n^0(\cdot,\underset{\sim}{Y}_1,\cdots,\underset{\sim}{Y}_n)$ is an admissible rule for the non-empirical problem for all $\underset{\sim}{Y}_1,\cdots,\underset{\sim}{Y}_n$ and n (see Houwelingen (1976), Meeden (1972)). However, to find such a sequence $\{d_{i,n}\}_{n=1}^{\infty}$ is very difficult, hence in view of Theorem 3.2.1, the more practical way to find the empirical Bayes rules is to estimate $\int_{x_i}^{\alpha_i} xc_i(x)dG_i(x)$.

<u>Lemma 3.4.1</u>. Let $p_i(x)$ and $G_i(x)$ be defined as above. If $m_{i,n}(x)$ is defined by (3.2.14) with $h \to 0$, $nh \to \infty$, then we have

$$\int_{x_i}^{\alpha_i} \frac{xp_i'(x)}{p_i^2(x)} m_{i,n}(x)dx - \int_{x_i}^{\alpha_i} \frac{x}{p_i(x)} dm_{i,n}(x) \to \int_{x_i}^{\alpha_i} xc_i(x)dG_i(x)$$

in (p).

<u>Proof</u>: Since

$$E\int_{x_i}^{\alpha_i} \frac{x}{p_i(x)}\, dm_{i,n}(x) = \int_{x_i}^{\alpha_i} \frac{x}{p_i(x)} \frac{1}{h}\,[m_i(x+h)-m_i(x)]dx$$

$$\to \int_{x_i}^{\alpha_i} \frac{x}{p_i(x)}\, dm_i(x) \quad \text{by LDCT,}$$

but

$$\mathrm{Var}\int_{x_i}^{\alpha_i} \frac{x}{p_i(x)}\, dm_{i,n}(x) = \mathrm{Var}\{\frac{1}{nh} \sum_{j=1}^{n} (U_j-V_j)\}, \text{ where}$$

$$U_j = \frac{Y_{ij}-h}{p_i(Y_{ij}-h)}\, I_{[x_i,\alpha_i]}(Y_{ij}-h), \text{ and}$$

$$V_j = \frac{Y_{ij}}{p_i(Y_{ij})}\, I_{[x_i,\alpha_i]}(Y_{ij}), \text{ hence}$$

$$\mathrm{Var}\int_{x_i}^{\alpha_i} \frac{x}{p_i(x)}\, dm_{i,n}(x) = \frac{1}{nh^2}\,\mathrm{Var}(U_1-V_1) \le \frac{1}{nh^2}\,E(U_1-V_1)^2$$

$$= \frac{1}{n}\int_{x_i+h}^{\alpha_i} [\frac{1}{h}(\frac{x}{p_i(x)} - \frac{x-h}{p_i(x-h)})]^2 dM_i(x) + \frac{1}{nh}\int_{\alpha_i}^{\alpha_i+h} \frac{1}{h}[\frac{x-h}{p_i(x-h)}]^2 dM_i(x)$$

$$+ \frac{1}{nh}\int_{x_i}^{x_i+h} \frac{x^2}{p_i^2(x)}\, dM_i(x)$$

$$\le \frac{1}{n} \max_{x\in[x_i,\alpha_i]} [\frac{d}{dx}\frac{x}{p_i(x)}]^2 + \frac{2}{nh} \max_{x\in[x_i,\alpha_i]} [\frac{x}{p_i(x)}]^2$$

$$\to 0 \quad \text{if} \quad nh \to \infty.$$

We see that

$$\int_{x_i}^{\alpha_i} \frac{x}{p_i(x)}\, dm_{i,n}(x) \to \int_{x_i}^{\alpha_i} \frac{x}{p_i(x)}\, dm_i(x) \quad \text{in (p).}$$

Similarly,

$$\int_{x_i}^{\alpha_i} \frac{xp_i'(x)}{p_i^2(x)} m_{i,n}(x)dx \to \int_{x_i}^{\alpha_i} \frac{xp_i'(x)}{p_i^2(x)} m_i(x)dx \quad \text{in} \quad (p).$$

Since

$$\int_{x_i}^{\alpha_i} xc_i(x)dG_i(x) = \int_{x_i}^{\alpha_i} - x \frac{d}{dx} [\frac{m_i(x)}{p_i(x)}]$$

$$= \int_{x_i}^{\alpha_i} \frac{xp_i'(x)}{p_i^2(x)} m_i(x)dx - \int_{x_i}^{\alpha_i} \frac{x}{p_i(x)} dm_i(x),$$

the proof is completed.

Now, let

$$\Delta_{i,n}^*(x_i) = \frac{\theta_0 m_{i,n}(x_i)}{p_i(x_i)} + \int_{x_i}^{\alpha_i} \frac{x}{p_i(x)} dm_{i,n}(x) - \int_{x_i}^{\alpha_i} \frac{xp_i'(x)}{p_i^2(x)} m_{i,n}(x)dx, \tag{3.4.1}$$

then

$$\delta_n^*(\underset{\sim}{x}) = \{i|x_i \geq \theta_0\} \cup \{i|x_i < \theta_0 \text{ and } \Delta_{i,n}^*(x_i) \leq 0\}$$

defines a sequence of empirical Bayes rules.

Empirical Bayes rules are useful only if we can control the rate of convergence. Johns and Van Ryzin (1971, 1972), Houwelingen (1973, 1976), Van Ryzin and Susarla (1977), and Gilliland and Hannan (1977) have derived theoretical upper bounds for $r_n(\underset{\sim}{G},\delta_n^*) - r(\underset{\sim}{G})$ under very general assumptions. Applying Lemma 3 of Van Ryzin and Susarla (1977), we get

Lemma 3.4.2. Let $\Delta_{G_i}(x) = \int_x^{\alpha_i} (\theta_0 - t)c_i(t)dG_i(t)I_{(0,\alpha_i)}(x)$, then

$$0 \le r_n(\underset{\sim}{G},\delta_n^*) - r(\underset{\sim}{G}) = \sum_{i=1}^{k} \{\int_{H_1} |\Delta_{G_i}(x)p_i(x)|P[\Delta_{i,n}^*(x) < 0]dx$$

$$+ \int_{H_2} |\Delta_{G_i}(x)p_i(x)|P[\Delta_{i,n}^*(x) \ge 0]dx\},$$

where $\Delta_{i,n}^*(x)$ and δ_n^* are defined by (3.4.1) and (3.4.2), respectively, and $H_1 = \{x|x < \theta_0$ and $\Delta_{G_i}(x) > 0\}$ and $H_2 = \{x|x \le \theta_0$ and $\Delta_{G_i}(x) < 0\}$. Now, let $O(\alpha_n)$ denote a quantity such that $0 \le \lim_{n\to\infty} \frac{O(\alpha_n)}{\alpha_n} < \infty$, Then since $|\Delta_{G_i}(x)p_i(x)| \le M_i$ for some constant M_i, so

$$r_n(\underset{\sim}{G},\delta_n^*) - r(\underset{\sim}{G}) \le \sum_{i=1}^{k} M_i\{\int_{H_1} P[\Delta_{i,n}^*(x) < 0]dx$$

$$+ \int_{H_2} P[\Delta_{i,n}^*(x) \ge 0]dx\}.$$

Therefore, if for all $x > 0$ and $n \to \infty$,

$$P[|\Delta_{i,n}^*(x)-\Delta_{G_i}(x)| > |\Delta_{G_i}(x)|] = O(\alpha_n),$$

then

$$r_n(\underset{\sim}{G},\delta_n^*)-r(\underset{\sim}{G}) = O(\alpha_n).$$

Now, by the inequality

$$P[|\Delta_{i,n}^*(x)-\Delta_{G_i}(x)| > |\Delta_{G_i}(x)|] \le \frac{Var[\Delta_{i,n}^*(x)]}{[|\Delta_{G_i}(x)|-|\Delta_{G_i}(x)-E\Delta_{i,n}^*(x)|]^2}$$

we get that if $Var[\Delta^*_{i,n}(x)] = O(\alpha_n)$ for all $x > 0$, then $r_n(\underset{\sim}{G},\delta^*_n)-r(\underset{\sim}{G}) = O(\alpha_n)$.

In the last part of this chapter, we let $X_i \sim U(0,\theta_i)$ for $i = 0,1$. θ_0 is treated as unknown. Assume that $g_i(\theta) = \frac{2\theta}{c^2} I_{(0,c)}(\theta)$ for $i = 0,1$ and $L_1 = L_2 = 1$. By Monte Carlo studies, we determine the smallest sample size N such that

$$\text{Relative error} = \frac{|r_m(\underset{\sim}{G},\delta^*_m)-r(\underset{\sim}{G})|}{r(\underset{\sim}{G})} \leq \varepsilon$$

for $N-4 \leq m \leq N$. The values of N corresponding to selected ε and c are shown in Table III.1 for $h = n^{-\frac{1}{4}}$, Table III.2 for $h = n^{-\frac{1}{5}}$, and Table III.3 for $h = n^{\frac{-1}{6}}$, where h is used to define (3.2.14).

TABLE III

Lists of values of the smallest N such that

$$\frac{|r_m(G, \delta_m^*) - r(G)|}{r(G)} \le \varepsilon \text{ for } N-4 \le m \le N,$$ where the

density of priors is $g_i(\theta) = \frac{2\theta}{c^2} I_{(0,c)}(\theta)$ for

$i = 0,1$.

$h = n^{-\frac{1}{4}}$

c \ ε	.25	.20	.15	.10	.05	.01
$\frac{1}{3}$	9	10	15	25	41	—
$\frac{1}{2}$	11	12	13	14	29	—
1	15	21	25	27	86	—
2	45	60	80	122	187	—
3	61	172	174	360	—	—

Note: "—" means that $N > 400$ (Monte Carlo study was curtailed because of limited resources).

TABLE III

(continued)

$h = n^{-\frac{1}{5}}$

$c \backslash \varepsilon$	.25	.20	.15	.10	.05	.01
$\frac{1}{3}$	11	13	15	21	27	—
$\frac{1}{2}$	10	13	15	21	48	—
1	13	19	20	21	46	—
2	26	27	52	151	262	—
3	51	88	134	232	304	—

TABLE III

(continued)

$h = n^{-\frac{1}{6}}$

$c \backslash \varepsilon$	.25	.20	.15	.10	.05	.01
$\frac{1}{3}$	9	10	15	25	41	—
$\frac{1}{2}$	11	12	13	14	29	—
1	11	15	20	27	97	—
2	19	31	59	60	212	—
3	51	61	136	171	302	—

Note: "—" means that $N > 400$ (Monte Carlo study was curtailed because of limited resources).

BIBLIOGRAPHY

Alam, K. (1973). On a multiple decision rule. Ann. Statist., 1, 750-755.

Bahadur, R. R. (1950). On the problem of theory of k populations. Ann. Math. Statist., 21, 362-375.

Bahadur, R. R., and Goodman, L. (1952). Impartial decision rules and sufficient statistics. Ann. Math. Statist., 23, 553-562.

Bahadur, R. R., and Robbins, H. (1950). The problem of the greater mean. Ann. Math. Statist., 21, 469-487. Correction. 22 (1951), 310.

Barr, D. R., and Rizvi, M. H. (1966). Ranking and selection problems of uniform distributions. Trabajos Estadist., 17, 15-31.

Bechhofer, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. Ann. Math. Statist., 25, 16-39.

Bechhofer, R. E., and Turnbull, B. N. (1974). A (k+1)-decision single-stage selection procedure for comparing k normal means with a fixed known standard: The case of common unknown variance. Tech. Report No. 242. Dept. of Operation Res., Cornell Univ., Ithaca, New York 14850.

Berger, R. (1977). Minimax, admissible and Γ-minimax multiple decision rules. Ph.D. Thesis. Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

Bhattacharyya, P. K. (1956). Comparison of the means of k normal populations. Calcutta Statist. Assoc. Bull., 7, 1-20.

Bhattacharyya, P. K. (1958). On the large sample behavior of a multiple-decision procedure. Calcutta Statist. Assoc. Bull., 8, 43-47.

Blum, J. R., and Rosenblatt, J. (1967). On partial a prior information in statistical inference. Ann. Math. Statist., 38, 1671-1678.

Carroll, R. J. and Gupta, S. S. (1977). On the probabilities of rankings of k populations with applications. J. Statist. Comp. Simul., 5, 145-157.

Carroll, R. J., Gupta, S. S., and Huang, D. Y. (1975). On selection procedures for the t best populations and some related problems. Comm. Statist., 4, 987-1008.

Deely, J. J. (1965). Multiple Decision Procedures from an Empirical Bayes Approach. Ph.D. Thesis (Mimeo. Ser. No. 45). Dept. of Statist., Purdue Univ., West Lafayette, Indiana 47907.

Deverman, J. N., and Gupta, S. S. (1969). On a selection procedure concerning the t best populations. Tech. Report. Sandia Laboratories, Livermore, California. Also Abstract. Ann. Math. Statist., 40, 1870.

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. J. Amer. Statist. Assoc., 50, 1096-1121.

Eaton, M. (1967). Some optimum properties of ranking procedures. Ann. Math. Statist., 38, 124-137.

Ferguson, T. S. (1967). Mathematical Statistics - A Decision-Theoretic Approach. Academic Press, N.Y.

Fox, R. J. (1978). Solutions to empirical Bayes squared error loss estimation problems. Ann. Statist., 6, 846-854.

Gilliland, D. C., and Hannan, J. (1977). Improved rates in the empirical Bayes monotone multiple decision problem with MLR family. Ann. Statist., 5, 516-522.

Gupta, S. S. (1956). On a decision rule for a problem in ranking means. Mimeo. Ser. No. 150. Inst. of Statist., Univ. of North Carolina, Chapel Hill, North Carolina.

Gupta, S. S. (1963). On a selection and ranking procedure for gamma populations. Ann. Inst. Statist. Math., 14, 199-216.

Gupta, S. S. (1965). On some multiple decision (selection and ranking) rules. Technometrics, 7, 225-245.

Gupta, S. S., and Hsu, J. C. (1978). On the performance of some selection procedures. Comm. Statist. B7(b), 561-591.

Gupta, S. S., and Huang, D. Y. (1975). On Γ-minimax classification procedures. Proceedings of the 40th Session of the International Statistical Institute, Vol. 46, Book 3, 330-335.

Gupta, S. S., and Huang, D. Y. (1977). On some Γ-minimax subset selection and multiple comparison procedures. Statistical Decision Theory and Related Topics (Ed. S. S. Gupta and D. S. Moore), Academic Press, New York, 139-155.

Gupta, S. S., and Panchapakesan, S. (1979). Multiple Decision Procedures: Theory and Methodology of Selection and Ranking Populations. John Wiley, New York. (To be published)

Gupta, S. S., and Sobel, M. (1958). On selecting a subset which contains all populations better than a control. Ann. Math. Statist., 29, 235-244.

van Houwelingen, J. C. (1973). On empirical Bayes rules for the continuous one-parameter exponential family. Ph.D. Thesis. Univ. of Utrecht.

van Houwelingen, J. C. (1976). Monotone empirical Bayes tests. Ann. Statist., 4, 981-989.

Hsu, J. C. (1977). On some decision theoretic contributions to the problem of subset selection. Ph.D. Thesis. Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

Huang, W. T. (1975). Bayes approach to a problem of partitioning k normal populations. Bull. Inst. Math. Acad. Sinica, 3, 87-97.

Johns, M. V., and van Ryzin, J. (1971). Convergence rates in empirical Bayes two-action problems. I. Discrete case. Ann. Math. Statist., 42, 1521-1539.

Johns, M. V., and van Ryzin, J. (1972). Convergence rates in empirical Bayes two-action problems. II. Continuous case. Ann. Math. Statist., 43, 934-947.

Karlin, S. (1968). Total positivity. Stanford University Press.

Lehmann, E. L. (1955). Ordered families of distributions. Ann. Math. Statist., 26, 399-419.

Lehmann, E. L. (1959). Testing Statistical Hypotheses. John Wiley, New York.

Lehmann, E. L. (1961). Some Model I problems of selection. Ann. Math. Statist., 32, 990-1012.

Lehmann, E. L. (1966). On a theorem of Bahadur and Goodman. Ann. Math. Statist., 37, 1-6.

Lehmann, E. L. (1975). Nonparametrics: statistical methods based on ranks. Holden-Day Press. San Francisco.

McDonald, G. C. (1974). The distribution of a variate based on indpendent ranges from a uniform population. Tech. Report GMR-1775, General Motors Research Laboratories, Warren, Michigan.

Meeden, G. (1972). Some admissible empirical Bayes procedures. Ann. Math. Statist., 43, 96-101.

Miescke, K. J. (1978). Bayesian subset selection for additive and linear loss functions. Mimeo. Ser. No. 78-23, Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

Miescke, K. J. (1979). Γ-minimax selection procedures in simultaneous testing problems. Mimeo. Ser. No. 79-1, Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

Mosteller, F. (1948). A k-sample slippage test for an extreme population. Ann. Math. Statist., 19, 58-65.

Paulson, E. (1949). A multiple decision procedure for certain problems in analysis of variance. Ann. Math. Statist., 20, 95-98.

Paulson, E. (1952). On the comparison of several experimental categories with a control. Ann. Math. Statist., 23, 239-246.

Puri, P. S., and Puri, M. L. (1969). Selection procedures based on ranks: Scale parameter case. Sankhyā Ser. A, 30, 291-302.

Randles, R. H., and Hollander, M. (1971). Γ-minimax selection procedures in treatment versus control problems. Ann. Math. Statist., 42, 330-341.

Rizvi, M. H., Sobel, M., and Woodworth, G. G. (1968). Nonparametric procedures for selecting a subset containing the population with the largest α-quantile. Ann. Math. Statist., 38, 1788-1803.

Robbins, H. (1951). Asymptotically subminimax solutions of compound statistical decision problems. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, Berkeley and Los Angeles, 131-148.

Robbins, H. (1955). An empirical Bayes approach to statistics. Proc. 3rd Berkeley Symp. Math. Stat. Prob. University of California Press, 155-163.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. Ann. Math. Statist., 35, 1-19.

van Ryzin, J. (1970). Empirical Bayes procedures for multiple decision problems. Tech. Report No. 249. Dept. of Statist., Univ. of Wisconsin, Madison.

van Ryzin, J., and Susarla, V. (1977). On the empirical Bayes approach to multiple decision problems. Ann. Statist., 5, 172-181.

Seeger, P. (1972). A note on partitioning a set of normal populations by their locations with respect to two controls. Ann. Math. Statist., 43, 2019-2023.

Singh, A. K. (1977). On slippage tests and multiple decision (selection and ranking) procedures. Ph.D. Thesis. Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

Tong, Y. L. (1969). On partitioning a set of normal populations by their locations with respect to a control. Ann. Math. Statist., 40, 1300-1324.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Mimeograph Series #79-16	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 6 Some Contributions to Gamma-Minimax and Empirical Bayes Selection Procedures.		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER Mimeo. Series #79-16
7. AUTHOR(s) 10 Ping Hsiao		8. CONTRACT OR GRANT NUMBER(s) [redacted] N00014-75-C-0455 15
9. PERFORMING ORGANIZATION NAME AND ADDRESS Purdue University Department of Statistics W. Lafayette, IN 47907	12 144 p.	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Washington, DC		12. ~~REPORT DATE~~ 11 August 1979
		13. NUMBER OF PAGES 140
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

14 MMS-79-16

18. SUPPLEMENTARY NOTES

9 Mimeograph series,

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Close to a control, Γ-minimax rule, Bayes rule, Minimax rule, t-Best population, PF-density, Better than a control, Empirical Bayes rule, Monte Carlo study.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The inadequacies of classical tests of homogeneity, in certain situations of interest to the experimenter, have long been known. Ranking and selection procedures which are formulated to overcome these shortcomings have provided realistic methods to handle the problems of comparing k populations. In this thesis, some results on Γ-minimax selection procedures and empirical Bayes selection procedure have been obtained. In Chapter I, a problem of selecting populations close to a control is considered. It is assumed that there are (k+1) normally distributed populations, and one of them is a control population.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601 | UNCLASSIFIED → next page

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

297 730

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

cont.

Our goal is to select those populations which are sufficiently close to the control. A zero-one type loss function is defined. Bayes rules, Γ-minimax rules and minimax rules are derived and compared. For Γ-minimax rules, some optimal properties are shown; also some general distributions are given for which Γ-minimax rules can be found.

Chapter II deals with the problem of selecting the t-best populations. It is shown that if the populations have Pólya Frequency Type II densities, then the natural selection rule is a Γ-minimax rule. This result has also been extended to the case where the populations are not necessarily independent. Also, by a simultaneous selection of the t-best populations for all $1 \le t \le k-1$, a Γ-minimax rule for complete ranking of the k populations is derived. Γ-minimax rules for some problems in testing hypotheses related to multinomial and multivariate negative binomial distributions are also derived.

In Chapter III, a problem of selecting populations better than a control is considered. When the populations are uniformly distributed, empirical Bayes selection rules are derived for a linear loss function for both the known control parameter and the unknown control parameter cases. When the priors are assumed to have bounded supports, empirical Bayes rules are derived for more general distributions. Monte Carlo studies are carried out which determine the minimum sample sizes needed to guarantee that the empirical Bayes rules are close to the true Bayes rule in terms of the Bayes risk.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)